(2)

IDA PAPER P-2427

IMPROVING THE CLASSIFICATION EFFICIENCY OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY THROUGH THE USE OF ALTERNATIVE TEST SELECTION INDICES

Cecil D. Johnson
Joseph Zeidner
Dora Scholarios

December 1990

Prepared for
Deputy Director of Defense Research and Engineering
(Research and Advanced Technology)

IDA

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract MDA 903 89 C 0003 for the Department of Defense. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

This Paper has been reviewed by IDA to assure that it meets high standards of thoroughness, objectivity, and appropriate analytical methodology and that the results, conclusions and recommendations are properly supported by the material presented.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1990

3. REPORT TYPE AND DATES COVERED
Final--February 1989 to November 1990

4. TITLE AND SUBTITLE
Improving the Classification Efficiency of the Armed Services Vocational Aptitude Battery Through the Use of Alternative Test Selection Indices

5. FUNDING NUMBERS
C - MDA 903 89 C 0003

T - T-D2-435

6. AUTHOR(S)
Cecil D. Johnson, Joseph Zeidner, Dora Scholarios

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-2427

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD(A)/R&AT
The Pentagon, Room 3D129
Washington, DC 20301

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study examines procedures for selecting tests for a battery as one means of improving the classification efficiency and potential economic benefit of the Army's current operational test battery. Five test selection methods, each representing competing measures of selection and classification efficiency, are compared under varying conditions expected to affect classification efficiency.

The research approach adopted involves a simulation of the Army selection and classification process based on the coverage of Army jobs provided by the Project A data. Comparisons of the classification efficiency produced under each test selection methodology are reported in terms of the average mean predicted performance (MPP) produced by each simulation.

The results confirm the predictions of Zeidner and Johnson (1989b) that the use of a classification efficient test selection procedure can improve the utility of the Army assignment process. The use of a method which maximizes potential classification efficiency results in as much as twenty percent gain in MPP over use of a method which maximizes predictive validity when used for a five test battery. In addition, doubling the number of jobs used in the simulation results in a performance gain of ten percent. This gain is an underestimate of what we would predict from increasing the number of classification efficient predictor composites or aptitude areas (AAs) corresponding to effectively reconstructed job families. Capitalization on disparate means and variances across jobs (hierarchical classification efficiency) provides no greater benefit for classification efficiency than is possible from a pure allocation efficient assignment strategy. This suggests that significant gains in productivity can be obtained simply by modifying the Army's current AA composites to capitalize on allocation efficiency, i.e., use the full set of least squares estimates of performance standardized to have equal means and variances.

Utility estimates of observed increases in MPP give an indication of the economic benefits of implementing classification efficient test selection methods. Gains of the magnitude found in this study could be worth well over 100 million dollars annually to the Army. Further gains may be expected when used in the context of different sets of predictors and jobs.

14. SUBJECT TERMS
selection classification, classification efficiency, allocation efficiency, selection utility

15. NUMBER OF PAGES
126

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39 18
298-102

IDA PAPER P-2427

IMPROVING THE CLASSIFICATION EFFICIENCY OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY THROUGH THE USE OF ALTERNATIVE TEST SELECTION INDICES

Cecil D. Johnson
Joseph Zeidner
Dora Scholarios

December 1990

IDA
INSTITUTE FOR DEFENSE ANALYSES
Contract MDA 903 89 C 0003
Task T-D2-435

ACKNOWLEDGMENTS

We are grateful for the reviews and useful technical comments made by Wayne S. Sellman, Military Manpower and Personnel Policy (Accession), Office of the Assistant Secretary of Defense; Richard C. Sorenson, Navy Personnel Research and Development Center; Leonard D. White, Army Research Institute; and Lauress L. Wise, Defense Manpower Center, Personnel Testing Division.

This study was performed for the Office of the Under Secretary of Defense for Acquisition (Research and Advanced Technology). Technical cognizance for the work, under Task Order T-D2-435, was assigned to the Assistant for Training and Personnel Technology. We are indebted to Earl Alluisi for his contributions to and interest in the effort, and to Jesse Orlansky, Institute for Defense Analyses, for suggesting that the utility of selection and classification procedures be evaluated, for providing valuable information, and for his many useful suggestions on preparing this report.

ABSTRACT

This study examines procedures for selecting tests for a battery as one means of improving the classification efficiency and potential economic benefit of the Army's current operational test battery. Five test selection methods, each representing competing measures of selection and classification efficiency, are compared under varying conditions expected to affect classification efficiency.

The research approach adopted involves a simulation of the Army selection and classification process based on the coverage of Army jobs provided by the Project A data. Comparisons of the classification efficiency produced under each test selection methodology are reported in terms of the average mean predicted performance (MPP) produced by each simulation.

The results confirm the predictions of Zeidner and Johnson (1989b) that the use of a classification efficient test selection procedure can improve the utility of the Army assignment process. The use of a method which maximizes potential classification efficiency results in as much as twenty percent gain in MPP over use of a method which maximizes predictive validity when used for a five-test battery. In addition, doubling the number of jobs used in the simulation results in a performance gain of ten percent. This gain is an underestimate of what we would predict from increasing the number of classification efficient predictor composites or aptitude areas (AAs) corresponding to effectively reconstructed job families. Capitalization on disparate means and variances across jobs (hierarchical classification efficiency) provides no greater benefit for classification efficiency than is possible from a pure allocation efficient assignment strategy. This suggests that significant gains in productivity can be obtained simply by modifying the Army's current AA composites to capitalize on allocation efficiency, i.e., use the full set of least squares estimates of performance standardized to have equal means and variances.

Utility estimates of observed increases in MPP give an indication of the economic benefits of implementing classification efficient test selection methods. Gains of the

magnitude found in this study could be worth well over 100 million dollars annually to the Army. Further gains may be expected when used in the context of different sets of predictors and jobs.

CONTENTS

TABLES

ABBREVIATIONS

AA	Aptitude Area
AE	Allocation Efficiency
AFQT	Armed Forces Qualification Test
AR	Arithmetic Reasoning
ASVAB	Armed Services Vocational Aptitude Battery
CE	Classification Efficiency
CMF	Career Management Family
CS	Clerical Speed
CTP	Core Technical Proficiency
DAT	Differential Assignment Theory
DV	Differential Validity
EI	Electronics Information
EPAS	Enlisted Personnel Allocation System
FLS	Full Least Squares
g	General Component
GS	General Science
H_a	Horst's Absolute Validity Index
H_d	Horst's Differential Validity Index
HCE	Hierarchical Classification Efficiency
JP-C	Joint Predictor-Criterion
LSE	Least Square Estimate

Max-PSE	Maximum Potential Selection Efficiency
MC	Mechanical Comprehension
MDS	Multidimensional Screening
MK	Mathematical Knowledge
Mod. H_d	Modified Horst's Differential Validity Index
Mod. PDI	Modified Point Distance Index
MOS	Military Occupational Specialty
MPP	Mean Predicted Performance
NO	Numerical Operations
PAE	Potential Allocation Efficiency
PC	Paragraph Comprehension
PC	Principal Component
PCE	Potential Classification Efficiency
PDI	Point Distance Index
PP	Predicted Performance
PSE	Potential Selection Efficiency
VE	General Verbal Ability
VG	Validity Generalization

SUMMARY

In earlier reports, major validation studies and meta- analytic studies were reviewed (Zeidner, 1987); technical issues in the development of decision theoretic utility models were traced (Zeidner and Johnson, 1989a); means of improving classification efficiency and its measurement were detailed (Johnson and Zeidner, 1990); and productivity gains attributable to varying job entry standards and assignment procedures using the Armed Services Vocational Aptitude Battery (ASVAB) were estimated (Zeidner and Johnson, 1989b). The present study examines alternative psychometric test selection indices as a means of improving the classification efficiency of the ASVAB.

Over the last two decades, the tests and test composites of the ASVAB have been selected to maximize predictive validity with little attention given to improving classification efficiency of the total set of tests in a multi-job, optimal assignment situation. Additionally, the number of tests per composite has been kept small and weights restricted to unity to simplify operational use of the composites. This emphasis on predictive validity and operational simplicity, required in a pre-computer age, can be shown to be fundamentally erroneous and outdated with respect to both empirical results and psychometric theory.

Although the amount of predictive validity provided by each test composite of a battery is a very poor indicator of classification efficiency, predictive validity has been used by some investigators as the basis of making potentially damaging recommendations concerning the ASVAB. For example, McLaughlin, Rossmeissl, Brandt, Wise, and Wing (1984) recommended a revised set of nine AAs even though the revised AAs showed an 18 percent reduction in differential validity; the investigators believed that such a reduction was an acceptable price to pay for a moderate increase in predictive validity. Hunter, Crosson and Friedman (1985) and Schmidt, Hunter and Larson (1988), relying completely on predictive validity to evaluate the effectiveness of test composites in a classification context, believed that a single test composite, a measure of general cognitive ability, would provide more classification efficiency than the existing ASVAB composites. Such a position in favor of a single measure is tantamount to claiming that the joint predictor-criterion space is unidimensional.

Our previous analysis showed that the current Army AA composites are of limited value, but we also show that considerable classification efficiency is potentially attainable from the present ASVAB if the battery is used in accordance with classification efficient procedures (Zeidner and Johnson, 1989b). We believe the ASVAB would possess even more potential classification efficiency (PCE) if its development had not been largely based on a search for increasing the predictive validity of specific aptitude tests rather than on procedures for increasing mean predicted performance (MPP).[1] Thus, we are not pessimistic regarding the future of tests developed for use in classification batteries. We acknowledge that PCE is difficult to achieve unless specific efforts are directed at developing predictors, identifying efficient tests for the battery and designing procedures that have as their goal increasing PCE.

The utility of an operational classification battery can be increased by: (1) improving the classification efficiency of the existing battery; (2) increasing the number of job families with their associated test composites within the existing job family structure; (3) substituting full least square (FLS) test composites for existing Army operational aptitude area (AA) composites; (4) introducing a single stage multidimensional screening (MDS) in place of the current two-staged system; (5) clustering jobs into job families by a method which maximizes classification efficiency. Aspects of the first and second way of increasing classification utility are examined by this research. Studies for evaluating all five strategies are described in Zeidner and Johnson (1989b, Chapter 5). All aspects of these five strategies not investigated in this study are currently being investigated in other studies based on Project A data.

A. OBJECTIVES OF STUDY

The primary purpose of this research is to estimate the benefits attainable from improving the efficiency of the Army's classification and initial assignment system through increased classification efficiency of the ASVAB. This is examined by contrasting test selection methods based on opposing personnel classification theories and estimating the utility that can be expected from the use of classification methods over selection efficient methods. The study is also designed to assess the effect of increasing the number of

---

1 PCE is defined as the MPP resulting from use of "best" assignment composites and an optimal assignment algorithm to place each entity in a job. Classification efficiency (CE) is defined as the MPP resulting from the use of actual assignment composites and actual assignment procedures to place each entity in a job.

classification efficient aptitude area (AA) predictor composites corresponding to job families and the effect of improving the hierarchical classification efficiency (HCE) of the assignment process. Each of these proposed changes to the current Army system is expected to improve classification efficiency.

Differential assignment theory predicts that classification utility can be maximized by the use of the full set of least squares estimates rather than the existing AAs as estimates of predicted performance. The changes in test selection and predictor/job family structure proposed in this study should provide a further increase in utility.

The tenets of differential assignment theory (DAT) can be contrasted to an approach which proposes predictive validity as the only relevant and useful means of achieving classification efficiency.[2] The existing Army AAs were designed to maximize predictive validity, and, as argued by Zeidner and Johnson (1989b, Chapter 1), fail to capitalize on differential validity as required for a classification battery. Zeidner and Johnson also introduced the distinction between hierarchical classification efficiency (HCE) and allocation efficiency as different but often overlapping sources of classification efficiency, and argue that neither are appropriately utilized by the current Army system. Maximizing differential validity improves potential allocation efficiency (PAE). The HCE component of potential classification efficiency (PCE) could be improved by capitalizing on divergent job validities or job importance and thus enabling differences in criterion performance across jobs to influence assignment decisions. In addition to a comparison of test selection methods, this study: (1) compares assignment strategies designed to represent pure allocation effects with those which also capitalize on HCE; (2) investigates the effect of number of jobs, or job families, to which personnel are assigned; and (3) investigates the effect of the number of tests in the operational battery. All of these effects are examined in terms of PCE.

Relatively cost-effective modifications in the way that aptitude information is used, in the context of the present Army job families and predictor composites, is shown to improve potential classification efficiency beyond what is possible under a selection efficient system as presently used. The research findings of this study show how the services could achieve increased productivity of a magnitude that would cost over 100 million dollars to achieve by use of higher selection standards.

---

2 Tenets of DAT include use of MPP for the measurement of classification efficiency; this theory is described in Chapter V.

One of the study's methodological goals is the determination of an appropriate psychometric index for selecting tests from an experimental battery with the goal of maximizing the classification efficiency of an operational test battery. Horst's index of differential efficiency (H_d) and modifications of H_d have been used in the limited number of studies designed to examine differential validity relative to selection efficient methods (e.g., Horst, 1954; Harris, 1967; McLaughlin et al., 1984). There is some evidence that H_d provides a less than perfect estimate of utility (e.g., Cronbach and Gleser, 1965; Johnson and Zeidner, 1990). H_d was shown to be an inaccurate measure of PCE when it has a large component of HCE (Johnson and Zeidner, 1990).

An alternative conception of the problem of maximizing differential validity was discussed by Johnson and Zeidner (1990) and is examined relative to H_d in this study. This alternative, the point distance index (PDI), is based on a geometric representation of the joint predictor-criterion space, where the sum of the differences in the multidimensional joint predictor-criterion space of performance estimates (LSEs) from the midpoint provides a measure of differential validity. This experiment compares the effect of H_d with this alternative psychometric index of test selection on PCE.

The study also provides one of the few demonstrations of a highly useful research approach: model sampling to create synthetic scores combined with a simulation of classification system features and the computation of mean predicted performance (MPP) as the final output after assignment. This MPP can easily be converted to dollars and entered into a consideration of trade-offs against costs in order to express the utility values of alternative operational procedures. The algorithms and software developed for this path-finding study provide most of the tools required to complete the other three of the four studies described by Zeidner and Johnson (1989b) in the Appendices of Chapter 5.

B. RESEARCH QUESTIONS

Zeidner and Johnson (1989b) describe a number of research issues and questions in Chapter 5. This study directly addresses several of these research questions and provides the software tools and practical experience with model sampling techniques to address the remaining research questions presented in Chapter 5.

The issues examined in this study are as follows:

(1) Can an emphasis on classification efficiency in selecting tests for an operational battery provide increased utility in the classification process? Which index is the best predictor of classification efficiency in an operational battery?

(2) How much increase in utility is provided through the doubling of the number of aptitude areas and the corresponding job families without considering the increased homogeneity within each family and the greater heterogeneity across families that should result from shredding out 9 job families into 18? Any gain in MPP in this study resulting from increasing the number of jobs from 9 to 18 is an underestimate of the gain that would result from a classification efficient reconstitution of job families to accomplish this increase.

(3) How much of the utility resulting from increased PCE can be produced by pure allocation effects as contrasted with both allocation and hierarchical classification effects?

(4) What effect would the reduction in the size of the operational battery from 10 to 5 have on utility in a classification situation. More generally, what effect does an increase in the number of tests in the battery have on PCE?

(5) Does the use of hands-on criteria reveal the presence of PCE to a greater extent than the use of rating criteria?

C. RESEARCH APPROACH

The research approach adopted in this experiment follows from the manpower allocation research of Sorenson and his colleagues in the 1960s (e.g., Sorenson, 1965; Niehl and Sorenson, 1968; Olson, Sorenson, Hayman and Abbe, 1969; Johnson and Sorenson, 1974). These researchers emphasized the need to simulate the essential processes of a selection and assignment system in order to accurately evaluate the actual and potential classification efficiency of the methods and tools used in alternative processes (e.g., McLaughlin et al., 1984; Johnson and Zeidner, 1990). Consideration of the complete context of assignment (i.e., the jobs, the applicant population, the predictor battery, the assignment variables, operational constraints and the allocation algorithm) permits a more accurate estimate of MPP by which to assess classification efficiency. It also avoids the need for complex mathematical formulations of the classification problem, thereby making it a more practical tool for estimating efficiency.

More specifically, the simulations conducted by these Army researchers used samples of synthetic entities rather than empirically collected measures. This technique of model sampling involves the transformation of normally distributed random numbers into synthetic test scores which have specific statistical characteristics in common with a prescribed population (Johnson and Sorenson, 1974). Used in place of empirical test data, synthetic entities have the benefits of a flexible sample size and known expected distribution, means and covariances. Using the concepts of factor analysis and matrix

algebra, entities can be generated and transformed to reflect experimental conditions and more closely approximate an applicant population than is possible using actual scores of empirical samples (Johnson and Zeidner, 1989b, Chapter 4).

The flexibility of the technique allows the evaluation of conditions which could affect the system but are not available in terms of actual empirical data. In particular, when used to evaluate the conditions which affect optimal classification, a model sampling experiment can eliminate the effect of correlated error between assignment and evaluation variables which, with empirical data, could only be controlled by seriously reducing sample size. Estimating the utility of alternative test selection methodologies or allocation algorithms is facilitated by simulating the rejection or assignment of entities to jobs. Synthetic scores, the input to the simulated model, are used to calculate the predicted performance scores (the "full least square (FLS) equation") for every entity in every job. The output of assignment by LSEs (MPP scores in standard score form) presents a basis of comparison for the classification efficiency of alternative situations. Thus, model sampling is an ideal approach for testing hypotheses related to optimal classification.

Model sampling is used in this study to carry out the proposed comparison of the PCE and PAE derived from alternative test selection methods. The process to be simulated is: (1) the rejection of the lower 30 percent on a continuum based on a variable having the characteristics of AFQT and (2) an optimal assignment procedure, based on a separate, full battery LSE for each job and a linear program algorithm which provides optimal assignment. Simulation conditions that are varied in this experiment include the number of jobs, the number of tests used in assignment composites and a choice of jobs based on the nature of the criterion. The inputs to the model are predicted performance scores (LSEs) for a sample of synthetic entities from test batteries constructed by five alternative methods. The outputs are MPP standard scores based on all the predictor variables and using the same weights across all assignment conditions. The aim is to analyze PCE through MPP, and, in so doing, identify the best index and conditions which affect optimality.

D. OPERATIONAL IMPLICATIONS OF FINDINGS

The experimental simulations conducted in this study suggest several areas for change in the operational use of the ASVAB. These can be summarized as follows:

(1) Classification efficient predictors, selected from the Project A experimental test pool by a method which maximizes differential validity, could augment the current operational battery. Differential assignment theory indicates that

practically significant increases in productivity can be obtained even with the addition of one or two classification efficient tests to the existing ASVAB battery. This study provides evidence that both Horst's index of differential efficiency (H_d) and the point distance index (PDI), when used to select these additional tests, will result in significantly more PCE than would result from the most efficient index of predictive validity (Max-PSE).

(2) Doubling the number of Project A jobs from nine to eighteen provides up to a ten percent increase in the overall estimate of performance. Subdividing (shredding) existing Army job families in order to double the number of composites (AAs) and corresponding job families would, if effectively done, increase the average validities of the composites and decrease the average intercorrelations among the LSEs based on the composites. Both of these effects would increase PCE beyond that obtained from increasing the number of jobs to which recruits can be assigned. It is expected that a more classification efficient method of forming 18 Army job families would exhibit considerably greater utility.

(3) Pure allocation effects, obtained by replacing the existing aptitude areas (AAs) with full least square (FLS) composites of predicted performance, account for most of the increase in PCE when classification efficient test selection methods are used. Hence, utility can be increased without the need for weighting assignment variables by the relative job validities or by values of performance on different jobs; i.e., without capitalizing on hierarchical layering effects.

(4) While it is commonly agreed that the predictive validity of a composite of four to six optimally weighted variables would not be substantially increased by addition of more predictors, it was found that classification efficiency increased by more than ten percent when the number of predictors in the optimally weighted composite was increased from five to ten.[3]

(5) Comparisons in this analysis were made among results obtained through the use of classification efficient, as compared to selection efficient, indices after the effects of selection (using a selection ratio of .70) were realized. If the gains were instead compared to those from existing ASVAB composites, the optimal set of classification efficient composites would be substantially greater in number.

---

3 Since all assignment composites used in this study utilized all the tests in the "battery" of selected tests, no distinction can be made between increasing the size of assignment composites and the size of the battery. Unfortuntely, we did not save the regression weights for the assignment composites and thus cannot say how many of the regression weights lie within some arbitrarily defined distance from zero; tests whose weights fall within such an interval could reasonably be considered as having been given a weight that is equivalent to deletion. We could infer that our results apply to both the size of assignment composites and the test battery--if the same proportion of tests from the 5 and 10 test batteries were to fall within such an interval.

(6) In most of the 5 test composites only three of these "best" tests come from the existing ASVAB; less than half of the "best" tests in the 10 test composites are in the ASVAB.

In a previous study (Zeidner and Johnson, 1989b), we estimated that implementing the tenets of differential assignment theory would bring about a large aggregate gain in MPP. Our "ball park" estimate of gains attributable to improved operational procedures to increase PCE exceeds 200 percent in the aggregate. We predicted that the largest contribution to PCE gains are FLS predictor composites; next are enlarged and restructured job families; and then the addition of classification efficient tests in the battery.

The change to FLS composites provides the maximum PCE available in the present ASVAB and job families. Prior results (Zeidner and Johnson, 1989b, p.64; Sorenson, 1965) estimated the gain over the existing operation to be about 100 percent. The present study showed an additional performance gain as high as 22 percent (or .07 of a standard deviation of MPP) using a PCE-efficient index over a predictive validity-efficient index to select five tests from an experimental battery of 29 tests. We find a non-trivial gain (of about 40 percent) in MPP by the use of FLS composites comprised of classification efficient tests over the present ASVAB FLS composites.[4] We know from the cost-benefit analysis of Nord and Schmitz (1989) that improvements of one- or two-tenths of a standard deviation may result in very large dollar gains to the Army each year.

---

4 We obtain this estimate (of 40 percent) by subtracting the gain due to selection from the MPP obtained in this study and comparing the resulting gain over random assignment with the comparable gain over random assignment shown by ASVAB FLS composites reported by Nord and Schmitz (1989).

I. RESEARCH ISSUES

A. IMPROVING CLASSIFICATION EFFICIENCY

Two erroneous schools of thought concerning the improvement of classification efficiency have contributed to the steady decline of classification efficiency in the aptitude areas (AAs) of the services. One group seeks to improve the set of test composites by eliminating overlapping tests and to otherwise reduce the intercorrelations among the test composites (e.g., Thorndike, 1950). The other group claims that classification efficiency need not be considered since maximizing predictive validity will also maximize classification efficiency (e.g., Hunter, Crosson and Friedman, 1985; Schmidt, Hunter and Dunn, 1987). In fact, whether the objective is to maximize the selection or classification efficiency of a fixed operational battery, a least square estimate of job performance based on *all* tests in the battery is the best predictor composite to represent a job or job family. When AAs consisting of less than the total number of tests in the battery are to be utilized, composites which maximize predictive validity (i.e., selection efficiency) are not the same ones as those which maximize classification efficiency.

Thus, the maximization of classification efficiency involves much more than the reduction of intercorrelations among test composites. Also, classification efficiency can never be improved by selectively eliminating tests from composites. Nor can classification efficiency be achieved by maximizing predictive validity without considering classification efficiency.

The potential classification efficiency (PCE) of an operational classification battery can be measured as the mean predicted performance (MPP) resulting from optimal assignment of personnel using least square estimates of performance (full least squares (FLS) composites) as the classification variables (Johnson and Zeidner, 1990). PCE can be improved or degraded by the test selection process utilized to make changes in the operational battery. It cannot be maximized by changing AAs, nor, in practical situations, by eliminating tests from AAs. However, whenever appropriate research data is collected on an experimental battery of predictors, one may improve the classification efficiency of AAs already maximized as FLS estimates by improving the PCE of the battery. Project A

provides such an opportunity. Except for the possible effects of Zeidner and Johnson's reports on selection and classification utility and the example provided by the present IDA project, the consideration of new predictors for inclusion in the ASVAB would most likely, like all other test selection by the services during the past two decades, have been essentially based on the severely limiting goal of improving predictive validity.

Maximum classification efficiency is obtained when each job has its own FLS aptitude area composite. The number of jobs in each job family would ideally contain the maximum number which permits reasonably large validation samples. The trend in the services has definitely been towards smaller and smaller numbers of job families forcing each AA composite to represent more heterogeneous sets of jobs. The Air Force has four job families and the Marine Corps has six. The Army has nine and the Navy has eleven, but both are under considerable pressure to conform to the other services. This trend, like the reliance on predictive validity to obtain PCE, is largely due to the impact of the validity generalization partisans' preference for believing that a single general aptitude measure is adequate for all operational purposes.

Zeidner and Johnson (1989b, Chapter 1) explain all classification effects as either hierarchical layering effects, allocation effects, or an inseparable combination of both. Hierarchical classification efficiency (HCE) can be obtained from capitalizing on disparate means and variances of predicted performance (PP) measures across criterion variables by weighting jobs according to their validities and/or value. Allocation efficiency (AE) can be achieved from capitalizing on differential validity; the differences in predicted performance across jobs within individuals. The current Army aptitude area (AA) composite predictors reflect no HCE and a disappointingly low amount of AE.[5] These Army composites were standardized to have equal means and variances and, since they are unit-weighted composites of three or more ASVAB tests and are not weighted by either job validity or value, they, consequently, possess no HCE. Furthermore, the ASVAB battery and its composites were not constructed to capitalize on differential validity and hence severely limit AE. The present ASVAB, however, has much more AE than can be obtained from using the existing AAs in making optimal assignments. This is a direct result of focusing on the predictive validity of the composites rather than on their classification efficiency.

---

5 It can be argued that since the Army appears to set higher cutting scores for entry into MOS with higher validity and/or greater value (in addition to the consideration of the "difficulty" factor), some HCE is indirectly introduced.

Zeidner and Johnson's (1989b) analysis showed that the Army AA composites hold considerable PCE if the ASVAB is used in accordance with differential assignment theory. This would imply the use of full least square (FLS) composites in place of the AAs as estimates of predicted performance which would provide all of the PAE present in the ASVAB. Weighting these FLS composites by job validities or values, depending on the availability of explicit statements of value from policymakers, would further provide HCE for augmentation of AE.

The present research assesses the effects of several of Zeidner and Johnson's (1989b) proposed changes in the operational use of the ASVAB which are intended to improve classification efficiency. Its focus is on the effect that test batteries selected to have inherently more PCE have on mean predicted performance after assignment; that is, on classification efficiency. The research design also permits examination of the effect on classification of having more AAs to be used for making assignments when the augmented set of assignment composites has not been improved with respect to homogeneity within job families or heterogeneity across composites. If there is a gain from having more composites under these conditions, one could conclude that only the "tip of the iceberg" has been observed with respect to the gains that might be expected from a classification efficient reconstitution of the Army AAs into a larger set of AAs and corresponding job families.

The effect of reducing the classification battery from 10 to 5 tests, selected either for maximization of classification or selection efficiency, is examined in this study. The effect of using "best" assignment variables with all HCE removed is also examined. The nature of the selected tests is considered and compared with those that have been included in the ASVAB. Finally, the implications of this study regarding how tests should be selected in order to improve the PCE of the ASVAB are presented.

A new index, the Point Distance Index (PDI), which was first reported in Johnson and Zeidner (1990), is compared to Horst's index of differential efficiency (H_d) with regard to its contribution to the PCE of classification batteries. Another index described in Johnson and Zeidner (1990), Max-PSE, is believed to be a superior measure of selection efficiency than Horst's index of absolute validity (H_a), although this is not empirically proven in this study. A comparison between H_a and Max-PSE is not made here since a simulation of selection and a measure of selection efficiency would be required. This is beyond the scope of the study.

B. ESTIMATING CLASSIFICATION EFFICIENCY

Constructing a test battery which maximizes classification efficiency implies the ability to measure the potential contribution of a predictor to the classification efficiency of the entire test battery (PCE) before it is selected for the battery. Horst (1954, 1956), who, along with Brogden, pioneered the methodology for constructing classification efficient batteries, stated the problem as one of defining an index of differential validity (DV) to "select those particular tests, no greater in number than prespecified, which would most accurately predict differences in success for all possible pairs of activities" (Horst, 1954, page 2). Horst's index of differential validity, H_d, only applies to the measurement of the PCE of a battery when FLS estimates are used for all assignment variables; H_d cannot be used when the assignment variables are to assess the classification efficiency of AAs. H_d could be modified to compare the capacity of alternative sets of test composites for capturing the potential DV, and hence PCE, of each such set of assignment variables (McLaughlin, et al., 1984; Johnson and Zeidner, 1990).

A battery constructed for selection decisions would favor predictors which provide the maximum average correlation with the separate criteria, also taking into account the effects of all previously selected predictors. To this end, Horst (1955) presented an index of absolute prediction (H_a) which, when used for test selection, would enhance but not maximize the battery's potential selection efficiency (PSE).

The index H_a is based on the average squared multiple correlation coefficient ($\mathbf{R}^2$) depicting the relationship between the test battery and each of the job criteria. This averaging of the $\mathbf{R}^2$ values instead of the $\mathbf{R}$ values makes H_a comparable to H_d, which is also a "squared" concept.

The multiple predictors of criterion success required by classification imply that both the criterion correlation coefficients and intercorrelations of predicted performance measures should be accounted for in any index of efficiency (Horst, 1954; Horst and MacEwan, 1960). Horst's (1954) index of differential predictive efficiency (H_d) was first conceptualized as the sum of the covariances describing the relationship between the differences among pairs of predictors and the differences among pairs of criteria. Horst provided an equivalent expression which made this covariance statistic more convenient for use as a test selection index. In this expression, H_d was defined in terms of the sum of the squared validities of the independent component of each trial test as a predictor of the differences among criterion scores. Computationally, this can be expressed as the sum of the variances of the difference scores between all possible pairs of predicted performance

measures. The larger the average variance as shown by H_d, the greater the differential prediction of the battery.

The value of test selection by H_d for classification was shown in an assignment simulation by Harris (1967). This compared H_d with Horst's index for absolute prediction (H_a) and showed that when a battery is to be used for optimal assignment, predictor selection by H_d results in greater classification efficiency than selection by H_a. Harris showed a ten percent improvement in the MPP of five tests selected by H_d over H_a.

Both the usefulness and sufficiency of H_d, however, have been debated. Cronbach and Gleser (1965) questioned H_d's applicability to assignment given that it does not consider rejected applicants, assuming instead that all individuals tested are to be assigned. They argued that H_d would only provide an efficient battery in "quota-free adaptive placement" where each individual's ability profile can be considered independently of external restrictions. In addition, the same authors questioned its accuracy by the fact that only in specific situations, such as equal criterion variances and weighting, will H_d be proportional to gains in utility.

This criticism is partially addressed in the present research which presents a modification of H_d as one of the five test selection indices to be compared. It is argued that this index (Mod. H_d) is a closer estimation of utility when efficiency cannot be derived from disparate predicted criterion variance (i.e., HC effects). In this situation, the efficiency index should be designed to detect the PAE inherent in the battery and in so doing provide a more accurate reflection of utility.

Recent research has compared H_d to Brogden's (1959) estimate of MPP. When all the classification efficiency in a battery is attributed to allocation effects (i.e., job validities and values are equal), Brogden's tabled statistic (an order statistic) times the product of the average absolute predictive validity of the LSEs (**R**) and the square root of one minus the average of the LSE intercorrelations is **r** PAE. Only in this situation is H_d, a measure of DV, proportional to Brogden's measure of MPP (Johnson and Zeidner, 1990, Chapter 2). This is the only direct link between H_d and the benefit component of utility that has been published. When HC effects contribute to PCE [i.e., job validities and predicted performance (PP) variances vary across jobs], there seems to be less justification for using H_d as an indicator of MPP (i.e., PCE). H_d contains variance sources that are not considered in Brogden's model, and Johnson and Zeidner (1990) show that these extraneous sources of variance cause H_d to vary away from MPP for several known conditions where both H_d and MPP can be computed. Under these circumstances, H_d does

not have the proven link to utility that Brogden's model can provide. Only the intuitive relationship described by Horst remains.

Horst's computational formula for H_d makes sense only when LSEs are used both as the measure of predicted performance and as assignment variables. Horst's intuitive concept, however, could be extended to the situation where predictors are not LSEs. Alternatives to H_d, while based on the same general principle, are designed to be used in less restrictive conditions. For instance, when the assignment variables are not LSEs as required by H_d, an alternative estimate of CE instead of PCE is necessary. As part of the Army's Project A, McLaughlin, et al. (1984) compared the DV obtainable from different sets of ASVAB aptitude area (AA) predictor composites for 98 jobs (MOS). They wished to treat each composite (including AAs that are not LSEs) as a predictor, in the sense of Horst's algorithm, to be assessed in the context of HC effects, highly correlated predictor composites, and a varying number of job families (m) across composite sets. They developed a modification of H_d which would permit, intuitively, a comparison of a Horst-like index of the efficiency of composites which are not LSEs to the H_d obtained using composites which are separate LSEs for each job. Specifically, differences between all possible criterion (MOS) pairs and differences between predictor pairs (the LSEs based on the AA composites corresponding to the MOS pair) were correlated to produce an index, M. One hundred percent efficiency was defined as the use of 98 LSEs to represent each composite so that there was one regression equation for each MOS. This was measured using the index H, where $mH^2 = H_d$ when jobs are unweighted. The "relative efficiency" of alternative composite sets as compared to this ideal was then computed as a ratio of M to H. This ratio was used to compare the different composite sets with each other. In one version of M and H, decisions involving MOS with high frequency accessions were given greater weight in the calculation of both M and H.

The results indicated that both the current and alternative sets of composites reflected only 45% to 67% of the DV contained in the ASVAB, with the DV generally being lower for higher frequency MOS. These results are more meaningful than would have been obtained using H or H_d for evaluating sets of composites that are not LSEs. However, they are not as meaningful results as could have been obtained using a more justifiable modification of H_d than is provided by M.

Horst's index of differential validity (DV) was based on his intuitive belief that "the higher the relationship between the differences between assignment variable scores for the *i*th and *j*th jobs with the differences between predicted performance (PP) for the *i*th and *j*th

jobs, as averaged over all values of *i* and *j*, the more effective the total set of assignment variables for making classification decisions." Horst provided a number of computing formulae which capitalized on his choice of the PP for each job as constituting *both* the set of assignment variables and the set of criterion variables. The computing formulae based on this assumption included the "squared differences" computing formula McLaughlin et al. chose to slightly modify to provide the index they call *H*. The use of the same PP variable as both the assignment and the criterion variable permits computing the DV index, either H_d or H, as a function of the squared differences between the PP scores of the *i*th and *j*th jobs. When the assignment variables are not PPs (e.g., as when the Army Aptitude Areas or experimental sets of classification composites are being evaluated), this simplification (the "squared differences" computing formula) is not available.

Unfortunately, the formula for *M* is based on a modification of the "squared differences" computing formula, despite the fact that *M* is intended for use when the assignment composite is *not* a PP composite, and for this reason alone is an inappropriate measure of DV. Also, instead of using the differences between the scores of the composites being evaluated as assignment variables as the predictors of the differences between corresponding criterion variables (what they should have done), McLaughlin et al. use the "best" weighted sum of the pair of test composites to predict criterion differences. The substitution of such a regression equation for the two assignment variables is certain to inflate the magnitude of the relationship between the differences of each pair of predictor variables and the differences of the corresponding pair of criterion variables. Thus, *M* provides an inflated estimate of DV which will be more inflated for some sets of composites than for others, as well as generally over-estimating the amount of total DV explained by operational and experimental composites in their role as assignment variables.

One of the central issues to be addressed in the present research involves a comparison of different test selection indices for maximizing PCE or PAE. Four alternative indices are compared to H_d. Firstly, an index aimed at maximizing P^^E (Max-PSE) is proposed as a more efficient alternative to Horst's index of absolute predic..on, H_a. Max-PSE, which is included to represent an assignment battery constructed to be selection efficient, provides a basis of comparison against the "PCE" indices (Johnson and Zeidner, 1990, Chapter 3). This maximizes the validity of the "best" single composite that can be obtained from a battery, where "best" applies to the prediction of criteria for a combined sample of all jobs. Max-PSE also maximizes the increment of selection efficiency of each successive test to be included as the next addition to a single "best" weighted composite to

be used for selection. A selected test is one which, together with the previous test(s) selected, provides the largest average multiple correlation coefficient.

Secondly, in order to eliminate the HC effects which make H_d lose its proportionality to PAE, the semi-partial correlation coefficients for each predictor are weighted to provide an adjustment which avoids the effects of unequal validities. This produces an adjusted H_d index which, it is hoped, will capture the PCE due to allocation efficiency previously masked by the effects of HC (Johnson and Zeidner, 1990, Chapter 3).

Thirdly, Johnson and Zeidner (1990) proposed a direct alternative to H_d. The point distance index (PDI) is more sensitive to the distribution of LSEs on different dimensions in the joint predictor-criterion space to provide comparable coverage of all jobs. Although two sets of jobs may be clustered differently in the joint space (implying a different degree of multidimensionality and PCE for each set), it is feasible that both sets result in the same sum of squared distances from the midpoint, and hence the same value of H_d. This implies that the more uneven the LSEs' coverage of some jobs as compared to others, the less effective H_d is as a measure of PCE in comparison to PDI.

The PDI is proposed as a correction for the above deficiency since it is more sensitive to the evenness of coverage of job LSEs. The greater the average distance among LSEs (as contrasted to the squared distances), the more even the coverage and the greater the potential for allocation efficiency. This index for test selection involves the maximization of an average distance from the midpoint of the multidimensional space. The distinction between PDI and H_d essentially lies in the methodological approaches they take. PDI uses the measurement approach of multidimensional scaling while the computation of H_d can be obtained as a factor analytic representation of selected tests as factors. PDI has been tentatively recommended as a better measure of PCE than H_d (Johnson and Zeidner, 1990, Chapter 3).

Finally, an adjusted PDI index is used in the same way and for the same purpose as the adjusted H_d index--to eliminate HC effects. The adjusted PDI is expected to be a better representation than the adjusted H_d of the allocation efficiency present in the battery.

In sum, the test selection comparisons draw from the original work of Horst which was directed at designing a predictor battery that maximized DV, and hence PCE. The addition or deletion of a test from a classification battery is accomplished using a figure of merit which optimizes DV or an alternative estimate of MPP (such as PDI). Improving DV

certainly increases the PAE and probably the HCE of a test battery; a multiplicative weighting of performance estimates according to their appropriate job validities will occur, taking advantage of HC effects. Several alternatives to Horst's H_a and H_d have been proposed as potentially better measures of PSE and DV. This research is designed to compare some of these alternatives in the context of several other research questions, including whether the number of jobs, the size of the battery and the type of criterion have practical effects on classification utility.

II. RESEARCH METHOD

A. THE MODEL SAMPLING APPROACH

A model sampling approach to personnel research allows the generation of random samples of synthetic test scores for simulating the selection and classification of any given population of applicants. At the cost of assuming knowledge of the statistical parameters of a population (e.g., the youth population), model sampling techniques allow as many and as large samples as desired. The parameters of each sample provide best estimates of the universe parameters.

Considering the need to accomplish a complex simulation of the classification process in the desired experiments, the use of a large data bank containing an unbiased empirical sample of the required test and criterion scores is the only alternative to the use of model sampling. Nord and Schmitz' (1989) simulation of alternative Army assignment strategies provides an example of the use of empirical samples of measures. In dealing with existing scores, they were obliged to make several assumptions regarding such parameters as: (1) the distribution of the youth population of scores (since the scores were from selected and assigned soldiers) and (2) the selection ratio (only selection ratios larger than were present when the data were collected could be simulated). The complexities of policy and management constraints required that individuals be used in much the same way as synthetic score vectors. Synthetic entities allow more thorough knowledge of and hence control over the samples. However, assuming that universe parameter estimates are credible and that the data bank is both large and unbiased, there are still advantages accruing to each of these alternative approaches.

Clearly, empirical scores derived from an unbiased data bank are more realistic with respect to the period in which the data was collected. Despite greater realism, however, such a simulation would represent only a cross section of the system in time and may not extend to the input coming in under changed conditions and policies. Simulated input based on a normal distribution of synthetic scores, followed by the imposition of policy constraints and decisions, may provide a better basis for extrapolating to future periods than imperfect samples drawn from a past era.

Model sampling as a source of test and criterion scores permits the simulation of input outside the range of the samples included in the data bank and permits unlimited sample sizes. Special samples can be drawn representing segments of the population that do not exist in the data bank. This is especially valuable when simulating and evaluating alternative selection policies. In this study, the primary advantage provided by the model sampling approach is that a realistic population can be designated, another analysis sample created and further large numbers of replications for 30 separate combinations of experimental conditions can be generated, each replication using a random sample of synthetic scores.

The synthetic scores produced by model sampling, with their prescribed expectations for variance-covariance matrices, are most readily generated with Gaussian (normal) distributions and selection effects imposed as a simple truncation, leaving the expected distribution of the upper portion of the sample with a truncated normal distribution. The applicant group being simulated would in fact be selectively self-censored in the upper tail of the bell shaped curve. That is, the higher the cognitive ability test scores, the less likely an individual is to enlist into the Army. This self-censoring effect can, at some additional cost, be duplicated in a model sampling experiment, but so far no other investigator has seen the necessity of achieving this degree of realism, and neither do we for our present research purposes.

The model sampling approach used by Harris (1967) to evaluate test batteries selected either by H_d or H_a provides a simpler precursor to the present study. The inputs to the simulation were the test scores for synthetic entities representing sets of selected tests (batteries) selected by the different methods. This approach generated scores based on the different batteries which would have the same statistical characteristics and covariance matrices as samples from a normally distributed youth population. Optimal assignment was again based on separate FLS composites for each job, and uniform quotas were assumed across jobs. In this way, the efficiency of differential over absolute assignment was established.

The present study has the advantage of being able to use the Project A pool of experimental predictor and criterion measures. Project A provides the best experimental test battery evaluated by the Army in over a decade and provides carefully designed and collected performance criteria. Input into 19 MOS spread over separate career management families (CMF) (representing about half of such CMFs) is included. By correcting Project A empirical validation results for selection and classification effects back to a youth

population, credible population parameters appropriate for these utility analyses are obtainable. The availability of the Project A data provides a rare opportunity to conduct a study of this nature and at this level of experimental rigor.

B. RESEARCH DESIGN

The study is designed to compare the effectiveness of five test selection methods in the context of three other factors affecting classification efficiency--the number of tests in the battery, the number of jobs to which entities are optimally assigned and the nature of the job sample.

The five test selection indices are: (1) Horst's differential index (H_d) which approximately maximizes differential validity; (2) the point distance index (PDI), which is an alternative to H_d aimed at providing an improved measure of PCE by favoring tests which provide a more uniform (hence increased multidimensional) coverage of the joint predictor--criterion space; (3) a modified Horst's differential index (Mod. H_d), which eliminates hierarchical classification effects and maximizes PAE; (4) a modified point distance index (Mod. PDI), which also eliminates hierarchical classification effects while improving the uniformity of coverage in the joint predictor--criterion space and maximizing PAE; (5) Max-PSE, which maximizes the predictive validity of the best single composite variable by choosing the combination of tests with the largest average multiple correlation.

A five-test battery and a ten-test battery are compared to assess the practical efficiency of the test selection methods. Given that the ten-test battery encompasses the five-test battery, it is assumed that the five selection treatments will be more differentiated in the smaller battery. This should provide greater sensitivity to methodological differences and any superiority of one index over another. However, if such differences can be observed under a larger battery, a stronger case can be made for the use of a classification efficient index. The ten-test battery should result in higher overall MPP because it is more representative of a full set of predictors.

A sample of nine jobs is compared to a sample of eighteen jobs.[6] An increase in MPP is expected when the number of jobs is increased from nine to eighteen.

---

[6] Of the 19 jobs for which validity data was available, the one with the smallest validation sample (N = 69) was dropped from the study. Dropping this one job permits us to have the same number of jobs in the two subsets of jobs (those with and those without hands-on criterion data).

Brogden's (1959) model depicted each job as providing a measurable expansion of the joint predictor-criterion space and thus enhancing classification efficiency. Brogden's model and assumptions are not replicated in this study. Therefore, the increase in MPP from nine to eighteen jobs should not be as large as expected by Brogden at the higher end of the distribution. Any observed increase in PCE or PAE when more FLS composites are used for selection would have implications for the design of validation studies for an organization having a large population of jobs such as that of the Army. Such an organization can usually provide further gains in MPP by constituting a larger number of job families so as to increase the homogeneity within job families and the heterogeneity across families and also increase $\mathbf{R}$ and decrease $\mathbf{r}$ in the function $\mathbf{R}(1\text{-}\mathbf{r})$. However, Brogden's multiplier of $\mathbf{R}(1\text{-}\mathbf{r})$ should be unaffected by the experimental conditions of this study; only his third multiplier, the order function, is varied.

A nine-job sample with hands-on criterion measures (e.g., work samples) is compared to a nine-job sample without hands-on measures (e.g., supervisory performance ratings). The more "objective" measures, which are necessarily more job-specific, are expected to increase dimensionality, and therefore enhance PCE. The eighteen job condition combines nine jobs from each criterion set.

Thirty possible treatment combinations result from each of the above independent variables (see Table 1). The relative effectiveness of one condition over another (i.e., each cell of Table 1) is established by comparing the results of a complete assignment simulation for every experimental condition. Each condition's selected test battery produces predicted performance scores (LSEs) for every individual in every job (assignment variables). The result of each assignment is the evaluation or dependent variable; that is, the mean predicted performance (MPP) for the assigned group is the unit of analysis that makes up the individual observations for use in the analysis of variance that permits the testing of statistical hypotheses. This MPP standard score is obtained by applying "best" weights, separately computed for each job using the designated population data, to the synthetic test scores of entities in each of the cross samples (after optimal assignment of entities)--and then averaging across all entities in each sample.

The cross-validation design employed in this study uses the designated universe to obtain weights for the predicted performance (PP) scores used for evaluation and an independent random sample from this universe to generate the weights to be used to compute the PP scores used in the cross-samples as assignment variables. For the purposes of this study, the "designated universe" represented by the Project A concurrent

validation data corrected for restriction in range and criterion attenuation to estimate predictor intercorrelations and validities for the youth population. Transformation matrices for the generation of the "analysis" sample are computed from the "designated population" covariance-validity matrices and applied to synthetic normal deviates to create samples of synthetic scores for each of the 18 MOS. These synthetic scores are generated for test and criterion variables. Predictor intercorrelations and validities for the creation of assignment variables are then derived from the synthetic test scores. The same number of entities were generated for each of the 18 "job" sub-samples as were available in the Project A empirical sample.

Table 1. Experimental Design

INDEP. VARS.	NO.TESTS →	5					10				
	SELECTION INDEX →	PDI	Hd	Mod PDI	Mod Hd	Max PSE	PDI	Hd	Mod PDI	Mod Hd	Max PSE
NO.JOBS ↓	CRITERION SOURCE ↓										
9	HANDS-ON TESTING										
	NO HANDS-ON TESTING										
18	9 WITH,9 WITHOUT HANDS-ON TESTING										

The entities for the 600 assignment replications (20 cross-samples under 30 experimental conditions) are generated in the same way as those in the analysis sample except that criterion variables are not generated; only test scores are used from the 20 cross-samples. However, the weights for assignment variables are computed from the "analysis" covariance-validity data. In the generation of cross-samples all entities in the lower 30% of the distribution of synthetic AFQT scores are rejected to form an Army input sample at a selection ratio of 0.70. Parameter values (weights) obtained from the "analysis" sample are applied to each cross-sample entity to compute the assignment variables (LSEs) and on the

basis of these scores, each entity is assigned to a job using an optimal assignment algorithm. The evaluation LSE scores, which provide a separate evaluation variable for each job, are averaged to provide the MPP score for the total assigned group of entities. The weights used to compute the evaluation LSEs for a given job are based on population data and are the same across all experimental conditions. Thus, correlated error between the assignment variables and the evaluation variable is avoided. This research paradigm is depicted in Figure 1.

Simulations of the classification process are conducted for 600 replications of synthetic scores of 198 entities in each. Each of these 600 sets of entities is one replication in a three-factor, repeated measures analysis of variance design. This design has the following factors and levels:

a. test selection method (5 levels: Horst's differential index (H_d), the Point Distance Index (PDI); modified H_d; modified PDI; and Max-PSE);

b. number of tests selected for operational battery (2 levels: 5 tests and 10 tests).

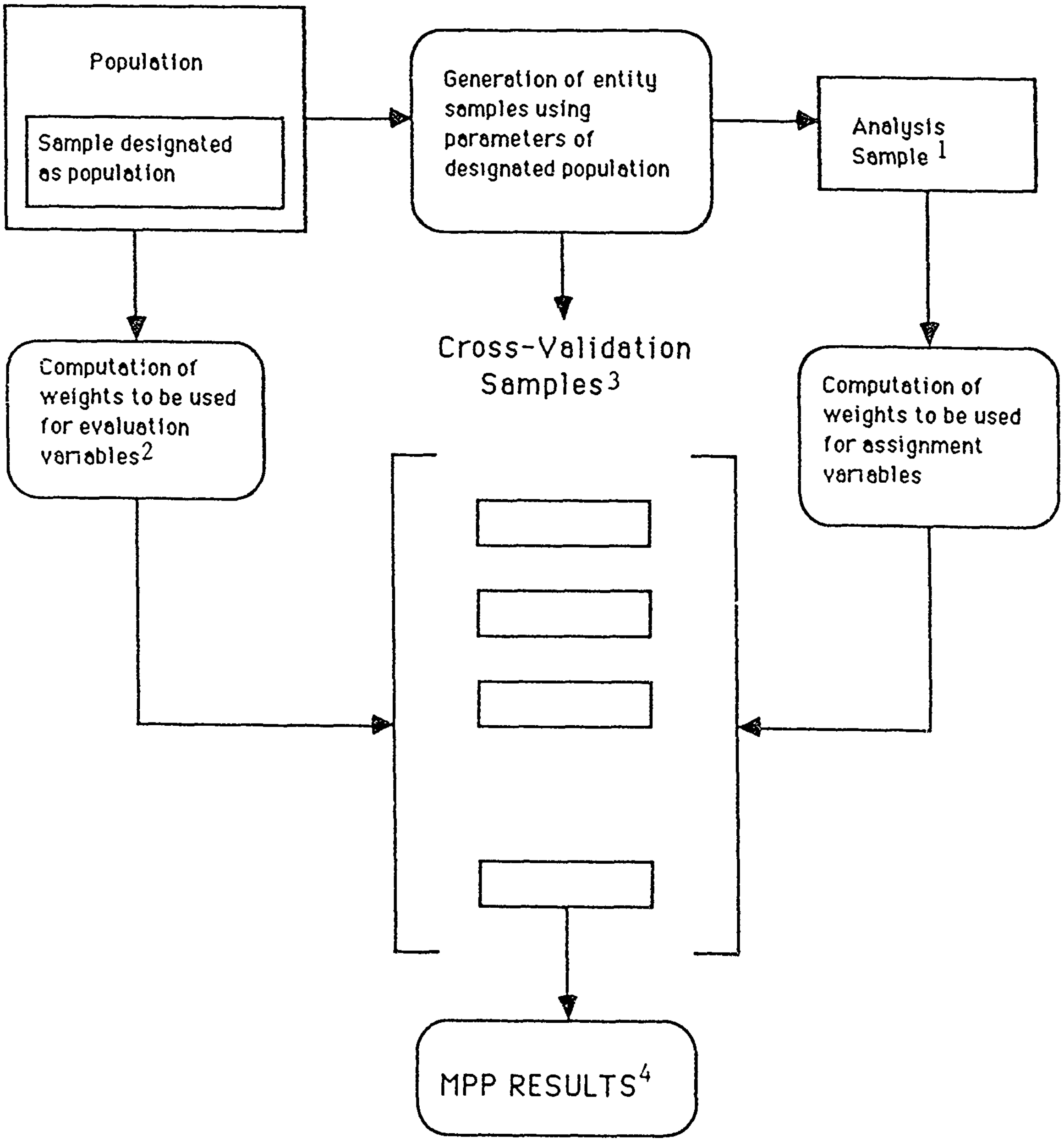
c. number and criterion type of jobs (3 levels: 9 jobs with hands-on criteria; 9 jobs without hands-on criteria; and 18 jobs with combined criteria).

Repeated measures (20 replications under each condition) are obtained for each of the 30 possible experimental conditions (i.e., for each cell of the results matrix in Table 1). The output of the simulation process, therefore, is an MPP standard score for each of the 600 replications of assigned entities.

C. PROCEDURE

1. Obtaining Youth Population Parameters from Project A Empirical Data

The empirical scores collected by Project A allowed the estimation of youth population parameters required for the experimental simulation. For 19 Military Occupational Specialties (MOS) (see Appendix A), Project A provided covariance data for 29 predictor variables (9 ASVAB tests and 20 additional experimental predictor constructs) and 5 criterion components. The experimental predictor measures were developed as potential additions to the current set of ASVAB AA predictor composites representing the 9 existing job families. These new predictors were designed to capture cognitive and non-cognitive abilities not covered by the ASVAB: spatial visualization and orientation, perception and psychomotor skills, temperament/personality, vocational interest, and job orientation (see Appendix B).



1 Job validation sample sizes equal to those used in Project A first-term concurrent validation study.

2 Evaluation weights computed from Project A empirical sample designated as the population.

3 Sample size of assigned entities number from 200-300; in the aggregate, N numbers in the thousands for each strategy.

4 Predicted performance is computed using the same evaluation variable and same weights for each job across all experimental conditions.

Figure 1. Typical Model Sampling Research

Five separate performance (criterion) components were identified in Project A as being descriptive of entry-level military job performance--general soldiering proficiency, core technical proficiency, effort and leadership, personal discipline and physical fitness and military bearing (Campbell, 1987; Wise, Campbell, McHenry and Hanser, 1986). Performance measures such as hands-on work samples, supervisory performance ratings, job knowledge tests and administrative measures were obtained from the concurrent validation phase of Project A (Wise, McHenry, Rossmeissl and Oppler, 1987).

Four of the criterion components--general soldiering proficiency, effort and leadership, personal discipline, and physical fitness and military bearing--play a role in performance across all Army MOS and are predicted by the AFQT composite and other selection indices, such as high school graduation (Wise, Campbell and Peterson, 1987). The fifth component, core technical proficiency, covers aspects of performance which are unique to each MOS. This proficiency is predicted by AA scores used for specific job families.

Two criterion formulations were available in the present study. In the first, a criterion composite was created from a weighted formula combining the five components according to their rated importance for each MOS. The criterion weights (Appendix C, Table C-3) were taken from Project A-initiated research by Sadacca, Campbell, White and DiFazio (1988) which examined the most appropriate method for criterion weighting. The second criterion formulation was represented by the core technical proficiency (CTP) component. Wise, et al. (1987) reported that the optimal component for differentiating between jobs on the basis of tests was CTP, with all other four components showing little added differential effectiveness in this respect. In the present study, the test selection results from each criterion were compared to determine which formulation provided greater differentiation.

Youth population parameters were generated from the empirical sample covariances for the 29 predictors, 5 criterion components and ASVAB intercorrelations for a national sample of American youths. These ASVAB intercorrelations (see Appendix D) represented the 1980 reference youth population for 18-23 year olds (Mitchell and Hanser, 1984).

An estimate of the youth population for the 29 predictors was obtained by implementing a data correction procedure for selection and classification effects. This algorithm encompassed two statistical corrections: a correction for restriction in the range of the sample due to selection and classification, and a correction for criterion unreliability.

The algorithm implementing these corrections was executed in the three stages summarized below.

a. Correction for Criterion Unreliability

A correction for measurement error is especially important in the case of the criterion composite since the reliabilities for criterion components are unequal. However, corrections were completed for both the composite and single criteria. All sample covariances involving at least one of the five criterion components were corrected for attenuation using the general formula with the uncorrected criterion component correlation in the numerator and the square root of the respective component reliabilities in the denominator. A reliability of 0.85 was used for general soldiering and CTP, and a reliability of 0.80 for effort and leadership, personal discipline and physical fitness and military bearing (Zeidner, 1987).

b. Correction for Selection Effects

The availability of the ASVAB youth population intercorrelation matrix enabled the ASVAB tests (the explicit selection variables) to be treated as variables drawn from an unrestricted population. Utilizing the predictor covariances for the total sample of incumbents, the remaining 20 restricted (implicit) predictors could be corrected for selection effects. The correction procedure was based on Lawley's (1943) assumption that the regression of the implicit predictors on the explicit predictors is linear and that the variance-covariance matrix of the latter exhibits homoscedasticity given the former. This correction provides an estimated variance-covariance matrix for the 20 implicit selection variables and the 9 by 20 matrix of covariances between the explicit and implicit correction variables. Gulliksen's formulae (1950, p.165, numbers 37 and 42) were applied to the youth population covariance matrix for the ASVAB tests (explicit variables) and the covariance matrix for the remaining 20 Project A predictors (implicit variables) for the aggregate of the 19 MOS. This produced a corrected variance-covariance matrix for all 29 predictor variables (but no criterion variables) representing the selected group.

c. Correction for Classification Effects

Using the same Gulliksen formulae as for the correction of selection effects, restriction of range corrections were completed for all covariances involving the 5 criterion components. Lord and Novick (1968) cite Lawley's (1943) proof for applying a second correction formula to an already corrected population. The corrected predictor covariance

matrix from the previous stage is used as the estimate of the population predictor covariances. At this stage all predictors are explicit selection variables and all criterion components are implicit selection variables. This procedure was conducted using first the 5 criterion components and then core technical proficiency (CTP) alone.

The result of the correction procedure was an estimate of the unrestricted (population) covariances with perfectly reliable criterion variables for the 19 samples. The corrected covariance matrix was converted to a matrix of intercorrelation coefficients among the 29 predictors. A weighted correlation of sums formula was then used to convert the validities against the 5 criterion components into a validity against a single MOS criterion variable. This produced unrestricted (i.e., universe) intercorrelations among the 29 predictors and two sets of the unrestricted (i.e., universe) validity coefficients--one for each of the 29 predictors against the composite criterion variable and another for each predictor against core technical proficiency.

2. Generation of Analysis Sample from Youth Population Parameters

The cross-validation design utilized random samples of synthetic entities representing the youth population that were generated through model sampling techniques. The empirical Project A data in turn was assumed to be an estimate of the universe predictor intercorrelations and validity coefficients based on perfectly reliable criteria, and was designated as the universe for the purposes of this study. The empirical sample designated as the universe and the generated samples differed only as a result of sampling error. Thus, operationally relevant decisions could be carried out on independently generated samples, each of which is assumed to be drawn from a population with known parameters (i.e., the designated universe).

The analysis sample had the same number of entities within each job as there were actual individuals in each MOS of the empirical sample. The comparability of the analysis sample and the sample designated as the universe was ensured by generating the synthetic scores such that the expected sampling errors for intercorrelation and validity coefficients approximated those of the Project A empirical sample.

The creation of the analysis sample proceeds from an **N** by **n** matrix of random normal deviates, where **N** is the number of entities and **n** is the number of tests. Employing the derivation of the basic structure of a correlation matrix (i.e., $\mathbf{R_t} = \mathbf{ADA'}$, where **A** is a matrix of eigenvectors and **D** is a diagonal matrix of eigenvalues), a Grammian factor solution of $\mathbf{R_t}$ can be obtained ($\mathbf{F_t} = \mathbf{AD^{1/2}A'}$). This is preferred over a

principal components (PC) factor solution since, in making more complete use of all predictors' random normal deviates, the normality of the distribution is improved.[7] Taking each of the 19 job samples individually, $\mathbf{F_t}$ can then be used to generate test scores for each entity in each job. The correlations of these scores with each other and with the criterion scores provide the intercorrelation matrix ($\mathbf{R_t}$) and the validity matrix (**V**) for the analysis sample. $\mathbf{R_t}$ and **V**, computed on the synthetic scores, have as their expected values the $\mathbf{R_t}$ and **V** of the designated population. This sample has the same expected variance-covariance matrix among the total set of predictors and criteria as does the corrected variance-covariance matrix computed from the Project A empirical sample. (See Appendix E, Section I for a full derivation of the analysis sample.)

3. Test Selection Procedure

Test selection was performed on the analysis sample predictor intercorrelation matrix ($\mathbf{R_t}$) bounded below by the analysis sample validities computed separately for all 29 predictors for each of the 18 Army MOS (matrix **V**). There were six versions of **V** (and hence $(\mathbf{R_t}|\mathbf{V}')'$) to represent the two criterion formulations and the three samples of jobs. For the composite and CTP criteria in turn, $\mathbf{V_{9A}}$ was created for the 9 jobs with hands-on criterion measures, $\mathbf{V_{9B}}$ for the 9 jobs without hands-on measures, and $\mathbf{V_{18}}$ for the combined set of 18 jobs. The same test selection algorithm was applied to all versions of $(\mathbf{R_t}|\mathbf{V}')'$.

Initially, the test selection routine was applied to the corrected, empirical Project A correlations (the designated universe) in order to determine which criterion formulation should be used in the analysis. The core technical proficiency (CTP) criterion displayed greater independence between the selected subtests of the different batteries than the composite criterion, and hence greater differentiation between job families. In addition, the overlap between selected tests appeared to be less for the "best" 5 predictor case than the "best" 10 predictor battery. This suggests that the use of the 5-test batteries against CTP

---

7 As compared to using a PC solution, where significant use is made of only the first few random normal deviates in the X matrix (XF = Y), all elements in each row of X contribute more equally to the value of each test score in Y when a Grammian factor solution is used as the transformation matrix, F. Thus using the Grammian factor solution, instead of the PC solution, provides the same benefit as using more random numbers to form each normal deviate--to better approximate a Gaussian distribution. Since we assume that our synthetic test scores are normally distributed, we felt that the use of both the Grammian factor solution and the sum of a number of random numbers to obtain each normal deviate was desirable.

criterion information would provide greater sensitivity for the detection of differing effects from the separate use of the five indices.

Although this result may seem anomalous given that the composite criterion resulted in more tests being selected overall (18 versus 15), the important finding for selecting one of the criteria for use in this study is the minimization of overlap between batteries used to represent the outcome of each of the five indices. In addition, the only other previous study with a bearing on the selection of tests to improve classification efficiency (Harris, 1967), used a criterion more comparable to CTP. Further, it is important that the utility results of the present study be comparable to those reported by Nord and Schmitz (1989). Nord and Schmitz also used a criterion comparable to CTP.

Wise, Campbell, and Peterson (1987) used LISREL, separately for each of the five criterion components, to test Project A predictor and criterion variables for unidimensionality in the joint predictor-criterion (JP-C) space. A subset of the 29 predictor variables of this study was used as dependent variables. The hypothesis that a single best weighted composite fits all jobs was rejected with respect to core technical proficiency; this hypothesis could not be rejected when the other four composites were each used as independent variables. The four Project A criterion components for which a single measure was an adequate predictor of performance across jobs should not be expected to have differential effects across jobs. The single task performance-based or "can do" measure among these four components was intended to reflect common military skills required of all soldiers, and the other three components, the motivationally based or "will do" measures, were intended to measure personal characteristics that we believe to be only slightly influenced by special job characteristics. The results of Wise et al. (1987) provide an additional argument for using the single criterion component designated as core technical proficiency as the criterion variable in this study.

Consequently, the decision was made to use the single (CTP) criterion results for all further analyses to determine the influence of different batteries of predictors on PCE and PAE. Having made this decision, the analysis sample $(\mathbf{R_t}|V')'$ matrix with the CTP criterion was used to select the five-test and ten-test batteries for use in the 30 experimental conditions.

The test selection algorithm after the first test is selected treats the orthogonal component of each predictor variable in the combined intercorrelation-validity matrix partly or completely as an orthogonal factor in the joint predictor-criterion space. By way of a

square root (triangular) factorization, a factor matrix (**F**) of orthogonal test components is constructed by successively adding a selected test as another column of **F**. The factoring procedure remains constant while the precise standard of merit (the independent variable test selection method) varies with the particular selection index used (i.e., H_d, PDI, Mod. H_d, Mod. PDI or Max-PSE).

The trial column of partial correlation coefficients providing the largest figure of merit becomes the next selected test and hence the next orthogonal factor to be added to **F** for each iteration. Using this accretion-type algorithm, ten tests are identified sequentially as having provided orthogonal components as columns of **F**. Execution of the algorithm was performed in FORTRAN for each condition. The programs were terminated when 10 tests had been selected in order to determine the "best" 5 and 10 tests.

The experimental test selection phase produced a total of thirty different batteries representing each of the cells of Table 1. The combined set of tests from all test batteries produced by the 30 experimental conditions is defined in this study as the full least squares (FLS) battery to be used in the calculation of evaluation weights and MPP scores for each assigned group. The full set of tests selected numbered twenty (see Appendix F, Table F-4).

4. Cross-sample Generation of Synthetic Scores

Based on the selected test batteries and the analysis sample predictor intercorrelations and validity coefficients, a total of 600 assignment samples with an N of 284 in each sample were generated--twenty such samples for each of 30 experimental conditions. These replications constituted the cross-samples on which assignment PP scores (LSEs) were computed using weights derived from the analysis sample. These weights were applied to the test scores generated using universe parameter values.

Predicted performance vectors for all entities in every job in each cross-sample have the same relationship to the values in the designated universe as PP vectors have to the same values computed in the analysis sample. This procedure can be summarized by the following four steps which are discussed below (matrix formula derivations of this procedure are given in Johnson and Zeidner (1990, Chapter 4)):

a. generation of random normal deviates;

b. transformation of normal deviates into test scores simulating the characteristics of the population of test scores from which the parameters are drawn;

c. transformation of test scores into predicted performance scores;

d. elimination of all performance scores below a cutting score on the AFQT score which is expected to eliminate 30% of all generated entities.

a. Generation of Random Normal Deviates

A uniformly distributed random sequence of numbers ranging from 0 to 1, with an approximate mean of 0.5 and a variance of 0.0833 was produced using a "purpose-built" pseudorandom number generator. The choice of a random number generator routine was based on evidence documenting efficient implementation and empirical tests of the randomness of the program's output (Park and Miller, 1988). A clearly defined algorithm, initial parameters and a recorded initial seed allow replication of the experiment. The optimal multipliers for producing the number sequence were based on Fishman and Moore's (1986) recommendations. Thus, potential defects in the random number generator that later generations may discern were minimized by careful selection of routines and inputs. By today's standards, the investigators were unusually cautious.

The sequence of uniformly distributed random numbers was transformed into a distribution of normal variables by calculating the expected mean and dividing by expected values to give a mean of 0 and standard deviation of 1.0. The vector of scores for each entity had the identity matrix as its expected covariance matrix, such that, $E(1/N(X_n'X_n)) = I_n$. This formed a matrix of normal deviates, $\mathbf{X_n}$, of order **N** by **n**, where **N** was the number of entities (individuals) and n was the number of simulated scores representing the full set of tests selected under all conditions. To generate evaluation, MPP, variables in this study, one sample of **N** = 284 and **n** = 20 was generated for each of twenty replications of the experiment. In generating assignment variables, the value of **n** changed according to the condition. Only five or ten tests, depending on the treatment, were used in obtaining the experimental LSEs.

b. Transformation of Random Normal Deviates to Test Scores

The generated test scores were required to have covariances with expectations equal to those of the score population being sampled by the selected tests and be represented by the comparable intercorrelations in $\mathbf{R_t}$, i.e., $E(1/N(Y'Y)) = \mathbf{R_a}$, or the cells of $\mathbf{R_t}$ corresponding to the selected tests. A Grammian factor solution ($\mathbf{F_t} = \mathbf{A}\mathbf{D}^{1/2}\mathbf{A}'$, where A and **D** were the eigenvectors and eigenvalues of the analysis sample predictor intercorrelations, respectively), was used to transform the matrix $\mathbf{X_n}$ to a matrix of test

scores (Y) for each of the experimental conditions. Given the transposed factor solution of R_a, $R_a = F_aF_a'$, Y was generated by $Y = X_nF_a'$, where F_a had dimensions **n** by **n** and Y 284 by **n**. Depending on the condition, **n** = 5 or 10.

c. Transformation of Test Scores to LSEs

For the set of **m** jobs (9 or 18) to be predicted by the **n** tests (5 or 10), the **N** by **m** matrix z of predicted performance scores was generated with an expected universe covariance matrix of $E(1/N(\mathbf{z}'\mathbf{z}))$. An **n** by **m** transformation matrix of beta weights, $W = R_a^{-1}V'$, was computed using the analysis sample data. This enabled the calculation of $\mathbf{Z} = \mathbf{YW}$. The **Z** for each set of jobs represented the **m** predicted performance scores of 284 simulated individuals and corresponded to LSEs of criterion performance for use as assignment variables.

A total of thirty 284 by m **Z** matrices were generated in accordance with the thirty experimental conditions. These thirty **Z** matrices depicted the product or performance of the five selection indices for two battery sizes (5 or 10 tests), against three samples of jobs (the 9-job, "objective" criterion set, the 9-job, "subjective" criterion set, and the 18-job set). Within each of the thirty cells of Table 1 twenty replications were executed. Thus, the output from this phase was 600 separate **Z** matrices of 284 by **m** representing LSEs as assignment variables with respect to every job for each artificial entity.

d. Selection of Entities by the AFQT Score

A selection ratio of 0.70 was accomplished by applying a truncation procedure to a ranking of the AFQT scores of entities within each sample. Within each replication, the entities with the lowest 30% of the AFQT scores were dropped from the analysis and not considered for assignment. The AFQT score was calculated by the formula combining the ASVAB tests Arithmetic Reasoning, Numerical Operations and Verbal Ability with equal weights given to the standard scores of each test.

5. Obtaining Performance Predictions from Assignment Simulations

For each replication, two assignment procedures are performed using either unmodified or modified LSE scores (the **Z** matrices) to capitalize on either HCE or PCE entirely free from HCE (i.e., PAE), respectively. In the first, the unmodified LSEs, with standard deviations equal to the validity, R_j for the jth job, are used as the assignment variables. The presence of hierarchical classification effects meant that the resulting MPP

standard scores reflect PCE with both HC and allocation effects. The second assignment procedure uses modified LSE scores divided by $\mathbf{R}_j$ as assignment variables. This standardizes the LSE scores to have equal expected means and variances across jobs and, since the hierarchical layering effect is eliminated, reflects a pure measure of PAE (i.e., PCE completely free of HC effects). Any classification efficiency evident in these MPP standard scores could be attributed to the allocation component of PCE.

In these two distinct assignment simulations, the modified and unmodified LSEs based on 5 or 10 tests are inputs, as assignment variables, to an iterative primal algorithm that is implemented by a circularized network optimization linear program.[8] In this algorithm, quotas are met at each iteration while the allocation sum converges toward the final optimal solution. At the final iteration the objective function, the MPP with respect to assignment variables, is maximized.

For the 9 and 18 job conditions quotas of 22 and 11 are used, respectively, so that all 198 selected entities are assigned. The final, optimal solution produces a set of associations between an entity and the job to which it is assigned (i.e., a 198 by 2 matrix with entity and job numbers in each row).

The MPP score for evaluation purposes is obtained from the assigned group in each experimental simulation. This is computed from the allocation information in the job-entity couplets given as the final solution. The 198 by 18 matrix of PP (LSE) scores is based on one set of regression weights derived from the universe rather than the analysis sample. These LSEs are generated using the full set of twenty predictors (the FLS equation) since this provides the best estimate of an entity's performance.[9] The MPP standard score for a given condition is obtained from the average of all the relevant evaluation FLS estimates across jobs and entities according to the order of assignment. This procedure is performed for each of the 600 cross-samples.

---

[8] The optimal assignment procedure was executed from the source code for the "NETG" mathematical programming system, published by Analysis, Research and Computation, Inc.

[9] Twenty predictors, instead of the total 29 predictors available from Project A, are used to provide estimates which hopefully reduce sampling error. The 9 predictors which were not included among the "best" 10 predictors by the use of any of the 5 indices were the only predictors which could be dropped from the set used to compute MPP without biasing the comparisons among indices. In general, the use of fewer tests to compute MPP favored Max-PSE.

III. EXPECTED FINDINGS AND ACTUAL RESULTS

An MPP standard score represents the average of expected performance for entities on the job to which each is assigned. The procedure described above produced two MPP standard scores for each of the thirty cells in Table 1 providing the total output of the system represented by that cell under "PCE" assignment and "PAE" assignment. Twenty replications and two separate assignments resulted in 600 measures of PCE and 600 measures of PAE. The expected relative magnitude of the MPP standard scores is expressed in the experimental statements or hypotheses I-VII belc ..

I. Overall efficiency of test selection indices

a. The magnitude of the MPP scores produced by the indices in the PCE assignment under each job condition and each battery condition will follow the hierarchical sequence PDI > H_d > (Mod. PDI> = Mod. H_d) > Max-PSE.

b. The magnitude of the MPP scores produced by the indices in the PAE assignment under each job condition and each battery condition will follow the hierarchical sequence Mod. PDI > Mod. H_d > (PDI> = H_d) > Max-PSE.

II. Classification efficient indices versus the selection efficient index

a. For the average of the PCE indices PDI and H_d combined across all conditions, the MPP will be significantly greater than the MPP produced by Max-PSE combined across all conditions.

b. The PCE MPP produced by PDI and H_d independently across all conditions will be significantly greater than the MPP produced by Max-PSE over all conditions. The relevant comparisons are between:

- H_d and Max-PSE;
- PDI and Max-PSE.

c. For the average of the PAE indices Mod. PDI and Mod. H_d combined across all conditions, the MPP will be significantly greater than the MPP produced by Max-PSE combined across all conditions.

d. The PAE MPP produced by Mod. PDI and Mod. H_d independently across all conditions will be significantly greater than the MPP produced by Max-PSE over all conditions. The relevant comparisons are between:

- Mod. PDI and Max-PSE;
- Mod. H_d and Max-PSE.

e. For each of the three job samples, the MPP (PCE and PAE) produced by PDI and H_d separately will be significantly greater than the MPP produced by Max-PSE for each job sample. The comparisons in hypothesis IIb and d are repeated for:

- the 9-job, "hands-on" criterion condition (9A);
- the 9-job, "no hands-on" criterion condition (9B);
- the 18 job, combined criterion condition.

III. H_d versus PDI for PCE and PAE

a. The PCE (measured as MPP) produced by PDI under all job and battery conditions will be significantly greater than that produced by H_d.

b. The PAE (measured as MPP) produced by Mod. PDI under all job and battery conditions will be significantly greater than that produced by Mod. H_d.

IV. Modified versus unmodified indices

a. The PAE (measured as MPP) produced by Mod. PDI under all job and battery conditions will be significantly greater than that produced by PDI.

b. The PAE (measured as MPP) produced by Mod. H_d under all job and battery conditions will be significantly greater than that produced by H_d.

c. The MPPs produced by the modified indices in the PAE-based assignment will be significantly greater than those produced in the PCE-based assignment.

V. Test battery size, MPP and differences in index efficiency

a. The MPP scores produced by the selection indices in the 10-test battery conditions will be greater than the MPPs in the 5-test battery conditions.

b. There will be greater sensitivity to differences in index efficiency when a 5-test battery is used than when a 10-test battery is used.

VI. Number of jobs and MPP

a. The PCE and PAE (measured as MPP) produced by the selection indices in the 18-job condition will be significantly greater than those in the 9-job condition.

b. The increase in the PCE and PAE (measured as MPP) between 9 and 18 jobs will approach but be smaller than the ratio representing the increase suggested by Brogden's (1959) model.

VII. **Source of criterion, MPP and index efficiency**

a. There will be greater sensitivity to differences in index efficiency under job sample 9A than under job sample 9B.

b. The MPP for PCE and PAE produced by the indices when hands-on criterion measures are used (job sample 9A) will be significantly greater than the MPP produced by no hands-on measures (job sample 9B).

The 600 MPP scores with and without hierarchical classification (the "PCE simulation" and the "PAE simulation"), can be organized under a three-way, $5 \times 2 \times 3$ analysis of variance (ANOVA) design, as shown in Table 2. This design encompasses repeated measures (i.e., 20 replications) under all factors.

Table 2. Three-Factor Analysis of Variance Design

Factor A: Selection Index
Level 1: PDI
Level 2: H_d
Level 3: Mod. PDI
Level 4: Mod. H_d
Level 5: Max-PSE
Factor B: Test Battery Size
Level 1: 5
Level 2: 10
Factor C: Job Sample
Level 1: 9 jobs/hands-on criteria (9A)
Level 2: 9 jobs/no hands-on criteria (9B)
Level 3: 18 jobs/mixed criteria

A null hypothesis of no difference between the treatment means was tested initially at the .01 confidence level to establish significant effects. Given the existence of a significant F-test of treatment means, more specific comparisons between means relevant to expected results I-VII were assessed. This involved a series of post hoc Scheffé multiple comparisons at the .05 level as a screening phase and more specific t-tests of pairwise comparisons between relevant MPP scores at the .01 level of significance.

A. TEST SELECTION RESULTS

The test selection procedure produced 30 batteries of five or ten tests. Table 3 below lists the tests and the conditions under which they were selected (i.e., the cells of Table 1).

Table 3. Test Selection Results By Selection Index, Test Battery Size, and Job Sample

SELECTION INDEX	NO. TESTS	JOB[a] SAMPLE	TESTS (IN ORDER SELECTED)[b]									
PDI	5	9A	MACH	CS	TECH	GS	AR					
	5	9B	AS	NO	PSYM	SPAT	MK					
	5	18	CS	AS	VE	MK	PSYM					
	10	9A	MACH	CS	TECH	GS	AR	AUTO	DEPN	NO	COND	MC
	10	9B	AS	NO	PSYM	SPAT	MK	CS	VE	SRAC	PSER	COMB
	10	18	CS	AS	VE	MK	PSYM	SPAT	NO	DEPN	COND	MACH
Hd	5	9A	MACH	CS	TECH	GS	AR					
	5	9B	AS	NO	SPAT	PSYM	MK					
	5	18	CS	AS	VE	SPAT	MK					
	10	9A	MACH	CS	TECH	GS	AR	AUTO	COND	DEPN	NO	MC
	10	9B	AS	NO	SPAT	PSYM	MK	VE	CS	SRAC	DEPN	GS
	10	18	CS	AS	VE	SPAT	MK	PSYM	NO	GS	DEPN	COND
MOD. PDI	5	9A	MACH	TECH	COMB	AUTO	DEPN					
	5	9B	DEPN	PSER	SRAC	COND	NO					
	5	18	MACH	DEPN	PSER	TECH	SRAC					
	10	9A	MACH	TECH	COMB	AUTO	DEPN	COND	CS	AR	GS	NO
	10	9B	DEPN	PSER	SRAC	COND	NO	MK	PSYM	SPAT	AS	CS
	10	18	MACH	DEPN	PSER	TECH	SRAC	NO	AS	PSYM	CS	SPAT
MOD. Hd	5	9A	MACH	COMB	TECH	AUTO	CS					
	5	9B	DEPN	VE	SPAT	MK	NO					
	5	18	MACH	CS	SPAT	MK	GS					
	10	9A	MACH	COMB	TECH	AUTO	CS	NO	MC	AR	GS	DEPN
	10	9B	DEPN	VE	SPAT	MK	NO	PSYM	AS	CS	SRAC	COND
	10	18	MACH	CS	SPAT	MK	GS	NO	PSYM	AS	DEPN	TECH
MAX-PSE	5	9A	AR	SPAT	EI	MACH	VE					
	5	9B	AR	GS	SPAT	CS	VE					
	5	18	AR	SPAT	GS	CS	AS					
	10	9A	AR	SPAT	EI	MACH	VE	DEPN	TECH	CS	CPAC	AUTO
	10	9B	AR	GS	SPAT	CS	VE	PSYM	AS	MK	PSER	DEPN
	10	18	AR	SPAT	GS	CS	AS	DEPN	CPAC	PSYM	NO	MK

a 9A: 9 jobs, hands-on criteria
9B: 9 jobs, without hands-on criteria
18: 18 jobs, mixed criteria

b Full test names are given in Appendix B.

The output from each test selection procedure was essentially a factor solution of the complete set of 29 predictors, with the order of factor presentation reflecting the sequence in which tests are selected (see Appendix F, section III). The key finding at this stage of the analysis is that different sets of predictors were selected using different selection indices. Across all conditions, a total of 20 tests were selected (see Appendix F, Table F-4 for FLS battery). Identical test batteries resulted from the PDI and H_d methods for three of the possible six combinations of battery size and job sample: 5/9A; 10/9A; and 5/9B. All other 27 test batteries were distinct.

Examination of the tests chosen by the PCE-oriented indices and Max-PSE supports the expectation that more tests highly loaded with general cognitive ability (e.g., Arithmetic Reasoning and General Science) would be selected by Max-PSE than by either the PCE or PAE indices. This provides preliminary but inconclusive support for the use of a specialized index to maximize CE. Also of interest is that a sizeable number of tests selected come from the experimental test battery rather than from the existing ASVAB. In addition, a difference between the modified and unmodified indices is apparent, suggesting that it may indeed be useful to distinguish between hierarchical classification (HC) and allocation effects as different sources of CE. However, analysis of assignment simulation results is required to estimate the impact of these different test batteries on PCE and PAE.

B. STATISTICAL ANALYSES OF SIMULATION RESULTS

All statistical analyses were performed on the MPP standard scores resulting from 20 replications of 30 experimental conditions. The overall F test (described below) was based on the results for the full 20 replications of each condition. Table 4 shows the average MPP for each assignment strategy reflecting each experimental condition, i.e., the MPP standard scores averaged across the 20 replications. The statistical tests for expected findings I-VII provide the basis for interpreting the aggregated figures in Table 4. The mean values in Table 4 that retain credibility after application of the statistical tests provide the basis for the utility analysis.

1. Results for Repeated Measures Analysis of Variance

The test batteries resulting from PDI and H_d under three of the conditions (see Table 2) represented essentially the same assignment treatment and so could not be treated as different conditions in the statistical analysis. A one-way ANOVA of the PDI and H_d

Table 4. Average MPP Standard Scores for 2 Assignment Strategies and 30 Conditions

Assignment With Hierarchical Classification ("PCE Assignment")										
	B_1					B_2				
	A_1	A_2	A_3	A_4	A_5	A_1	A_2	A_3	A_4	A_5
C_1	.38511	.38511	.35207	.37970	.31532	.44369	.44369	.44391	.45271	.42247
C_2	.39249	.39249	.33376	.35545	.35106	.47398	.46894	.46813	.46782	.45675
C_3	.39215	.40584	.36408	.40051	.39691	.49026	.46663	.48659	.50766	.48257
Assignment Without Hierarchical Classification ("PAE Assignment")										
	B_1					B_2				
	A_1	A_2	A_3	A_4	A_5	A_1	A_2	A_3	A_4	A_5
C_1	.37328	.37328	.33710	.35260	.29799	.43151	.43151	.43430	.44036	.40726
C_2	.38851	.38851	.32601	.35351	.31218	.45191	.45691	.45538	.45578	.44451
C_3	.35962	.36347	.34400	.37341	.36748	.47358	.44256	.46681	.48925	.46180

conditions which did differ was not significant at $p = .01$ ($F_{1,19} = 7.60$). Given that PDI and H_d appeared to be interchangeable even in the conditions where they could be distinguished, levels 1 and 2 of factor A were collapsed into one level to represent a single classification efficient index. Consequently, a $4 \times 2 \times 3$ analysis of main effects and interactions was initially used to analyze the overall differences between population means under each assignment. Tables 5 and 6 below display the results of these F tests.

The significant main effects of the three factors under both strategies allowed the null hypothesis of no difference between population means to be rejected at a high level of confidence (above .001). Significant interactions warn that the results of further multiple comparisons and t-tests of combined means cannot be interpreted simply as caused by the particular factor being evaluated. Hence, three levels of hypothesis testing were undertaken: (1) multiple comparisons using the Scheffé S method at $p = .05$, (2) t-tests between means combined over two or more relevant cells of the results matrix at $p = .01$, and (3) t-tests between two individual cells of the results matrix at $p = .01$. The results of these tests for expected findings or hypotheses I-VII are described below.

Table 5. Repeated Measures ANOVA of MPP Standard Scores: "PCE Assignment" (N = 480)

Source of variation	df	$MS_{treatment}$	MS_{error}	F	p
Between Subjects	1,19				
Subjects within groups					
Within Subjects					
A (Selection Methods)	3	0.019		82.15	<.001
A X subjects within groups	57		0.000231		
B (Test Battery Size)	1	1.115		2473.95	<.001
B X subjects within groups	19		0.000451		
C (Job Sample)	2	0.672		73.96	<.001
C X subjects within groups	38		0.009086		
AB	3	0.010		59.21	<.001
AB X subjects within groups	57		0.000169		
AC	6	0.007		24.47	<.001
AC X subjects within groups	114		0.000286		
BC	2	0.007		46.53	<.001
BC X subjects within groups	38		0.000150		
ABC	6	0.002		16.14	<.001
ABC X subjects within groups	114		0.000124		

Note: Due to rounding error, the F shown above is not exactly equal to the ratio of $MS_{treatment}$ to MS_{error}.

Table 6. Repeated Measures ANOVA of MPP Standard Scores "PAE Assignment" (N = 480)

Source of variation	df	$MS_{treatment}$	MS_{error}	F	p
Between Subjects	1,19				
Subjects within groups					
Within Subjects					
A (Selection Methods)	3	0.025		105.50	<.001
A X subjects within groups	57		0.000237		
B (Test Battery Size)	1	1.231		5083.84	<.001
B X subjects within groups	19		0.000242		
C (Job Sample)	2	0.039		49.92	<.001
C X subjects within groups	38		0.000781		
AB	3	0.011		60.01	<.001
AB X subjects within groups	57		0.000183		
AC	6	0.008		19.82	<.001
AC X subjects within groups	114		0.000401		
BC	2	0.005		26.84	<.001
BC X subjects within groups	38		0.000186		
ABC	6	0.002		14.20	<.001
ABC X subjects within groups	114		0.000141		

2. Results for Hypothesis I: Overall Index Efficiency

The MPP standard scores in Table 4 formed the basis for comparisons between means. In the "PCE assignment" strategy, examination of absolute differences in magnitude between selection methods ignored the distinction between PDI and H_d, although where they differed PDI seemed marginally superior.[10] The expected increase in magnitude for the CE index (the average of PDI and H_d) over Max-PSE was upheld in all conditions except for 10/18 (tests/jobs). When considered as separate methods, a reversal in the expected direction occurred with PDI in the 5/18 condition and with H_d in the 10/18 condition.

In the "PAE assignment" strategy, the important comparison was between Mod. PDI and Mod. H_d on the one hand and Max-PSE on the other. Both Mod. PDI and Mod. H_d produced greater MPP than Max-PSE in all conditions except for the 5/18 condition in which there was a reversal with Mod. PDI.

There appear to be no consistent increases in magnitude to support the expectation that the modified PDI or H_d indices produce greater MPP in a pure allocation efficient assignment strategy ("PAE assignment") than the unmodified PDI and H_d. However, without evaluating the statistical significance of these observed differences, no conclusions are justified at this stage.

3. Results for Hypothesis II: Classification Efficient Indices versus the Selection Efficient Index (Max-PSE)

The initial comparison between selection methods treated the average of PDI and H_d conditions as representing one classification efficient (CE) index. In the "PCE assignment" case, a t-test of the differences between CE and Max-PSE over all conditions was significant at the .01 level ($t_{19} = 11.321$), thus confirming the observed difference in magnitude. The means for PDI and H_d in the three conditions where they did differ (10/9B; 5/18; and 10/18), were compared with Max-PSE. This showed that the PDI mean at the 10/9B and 10/18 conditions was significantly different from the Max-PSE mean of the same conditions, but that the same could not be said for the H_d means (see Table 7).

---

[10] Although the null hypothesis could not be rejected at the selected level of significance ($p < .05$), the overall PDI results were superior to H_d, and particularly so for the replications representing the 9-job sample with hands-on criteria, there was a reversal in the single condition where 10 tests were selected and assignments were made to 18 jobs.

However, t-tests between PDI and Max-PSE ar H_d and Max-PSE in the 18 job conditions were not significant (see Table 7).

Table 7. t Values for Comparisons between PDI, H_d, and Max-PSE: "PCE Assignment"

Comparison	$B_2/C_{2,3}$	$C_3/B_{1,2}$
PDI v Max-PSE	4.4150*	− 0.383
H_d v Max-PSE	0.665	0.921

Note: PDI and H_d were different only in conditions B_2C_2, B_1C_2 and B_2C_3

* Significant at .01 level

In the "PCE" analysis, then, a combined CE index is superior to Max-PSE, but the consistency of this difference weakens when PDI and H_d are compared separately with Max-PSE. It is possible that the inherent weakness of the 10 test and 9B job conditions for detecting methodological differences [as predicted by hypotheses V(b) and VII] is responsible for these mixed results.

Some support for this position is found in the "PAE assignment" case where Mod. PDI and Mod. H_d could both be compared to Max-PSE means over all conditions. A Scheffé test indicated that significant differences existed between the three groups at the .05 level. In addition, t-tests between Mod. PDI and Max-PSE, and Mod. H_d and Max-PSE were significant at $p = .01$ ($t = 5.437$ and $t = 13.045$, respectively).

The superiority of the classification efficient methods over Max-PSE was also expected to carry across all possible job and test combinations [hypothesis II(e)]. A series of single factor ANOVAs was conducted examining the simple main effects of factor A (selection methods) for different combinations of B (test battery size) and C (job sample). The modified indices were dropped from factor A in the "PCE assignment," leaving an analysis of the simple effects of two levels of factor A (the combined PDI/H_d and Max-PSE). Similarly, the unmodified indices were dropped in the "PAE assignment" leaving three levels of A (Mod. PDI, Mod. H_d and Max-PSE). t-tests were also carried out, comparing the combination of PDI and H_d with Max-PSE in the "PCE assignment," and Mod. PDI and Mod. H_d each with Max-PSE in the "PAE assignment." The results of these analyses are summarized in Tables 8 and 9, respectively.

Table 8: Simple Main Effects and t Values of Selection Methods PDI/H_d and Max-PSE: "PCE Assignment"

Source of variation	$F_{1,19}$	t_{19}	p
A at B_1	166.49	−12.903	<.001
A at B_2	22.47	−4.740	<.001
A at C_1	159.37	−12.624	<.001
A at C_2	51.75	−7.194	<.001
A at C_3	0.06	0.254	n.s.
A at B_1,C_1	164.41	−12.822	<.001
A at B_1,C_2	57.25	7.566	<.001
A at B_1,C_3	0.17	−0.410	n.s.
A at B_2,C_1	41.28	−6.425	<.001
A at B_2,C_2	17.29	−4.158	<.001
A at B_2,C_3	1.03	1.013	n.s.

Note: n.s. signifies not significant at .01 level.

Table 9. Simple Main Effects and t Values of Selection Methods Mod. PDI, Mod. H_d Max-PSE: "PAE Assignment"

Source of variation	$F_{2,38}$[a]	t_{19}[b]	p	t_{19}[c]	p
A at B_1	65.69	−3.221	0.01	−11.137	<.001
A at B_2	42.66	−5.486	<.001	−9.179	<.001
A at C_1	135.60	−11.917	<.001	15.801	<.001
A at C_2	17.60	−2.785	0.01	−5.930	<.001
A at C_3	19.62	2.200	n.s.	−3.978	<.001
A at B_1,C_1	79.64	−8.770	<.001	−12 245	<.001
A at B_1,C_2	20.44	−2.102	n.s.	−6.281	<.001
A at B_1,C_3	18.25	4.560	<.001	1.153	n.s.
A at B_2,C_1	62.97	−8.609	<.001	−10.539	<.001
A at B_2,C_2	4.72	−2.612	n.s.	−2.707	n.s.
A at B_2,C_3	16.24	−0.977	n.s.	−5.350	<.001

Note: n.s. signifies not significant at .01 level.

a The numerator of the F ratio is the MS of A at each level of B and C; the denominator is $MS_{A\ x\ subj.\ within\ groups}$.

b Comparison between Mod. PDI and Max-PSE.

c Comparison between Mod. H_d and Max-PSE.

In conclusion, the composite CE index, created from PDI and H_d in the PCE condition, was found to be significantly greater than Max-PSE under all treatments excluding those with 18 jobs. In "PAE assignment," the superiority of Mod. PDI and Mod. H_d was most consistently shown for job sample 9A, both the 5- and 10-test conditions and any combinations of these three levels of factors B and C.

4. Results for Hypothesis III: H_d versus PDI

The distinction between H_d and PDI under "PCE assignment" could only be examined by statistical methods for three of the six possible combinations of job/test conditions: 10/9B; 5/18; and 10/18. In the other three conditions, any expected differences in the test batteries produced by H_d and PDI were disproven at the test selection stage. Further confirmation of the earlier non-significant F test for these two levels of factor A was given by a t-test between the PDI and H_d means at the possible conditions ($t_{19} = -2.757$, $p = .01$).

The distinction between Mod. PDI and Mod. H_d in the "PAE strategy" could be made for all relevant cells. A one-way ANOVA on levels 3 and 4 of factor A showed a significant effect at $p = .01$ ($F_{1,19} = 43.70$). Although this implies a rejection of the null hypothesis of no difference, examination of the MPP standard scores in Table 4 shows that Mod. H_d is consistently higher than Mod. PDI in direct contrast to the predicted direction. The expected increase as a result of using PDI in place of H_d is not supported among the modified indices.

The implication of these results, in combination with the finding reported for hypothesis III, is that either PDI or H_d may be used with high confidence that they are superior to Max-PSE.[11]

5. Results for Hypothesis IV: Modified versus Unmodified Indices

Hypotheses IV (a)-(c) predicted some benefit from the use of indices modified to eliminate hierarchical classification efficiency (HCE) in the pure allocation assignment strategy. Using the "PAE assignment" data, a one-way ANOVA for the PDI and Mod. PDI levels of factor A was carried out. This showed a significant effect of the index treatment at the .01 level ($F_{1,19} = 85.34$). A further t-test between the means over all

[11] The high statistical significance of this conclusion is obtained despite the reversal for one condition (10 tests and 18 jobs).

conditions was also significant ($t_{19} = -9.238$, $p = .01$). However, a similar ANOVA for the H_d and Mod. H_d levels of A in the same data was not significant at .05 ($F_{1,19} = 1.56$). These results provide no consistent evidence in support of the hypothesis that the proposed modification to selection methods would increase classification efficiency in conditions with no HCE.

6. Results for Hypothesis V: Test Battery Size, MPP and Index Efficiency

Hypothesis V(a) predicted that the larger the size of the test battery the higher the MPP resulting from assignment. An examination of Table 4 indicates an average increase of 28% in MPP from the 5-test to the 10-test conditions in the "PCE assignment," and of 24% in the "PAE assignment." The difference between the total group of 5-test means and the total group of 10-test means was also statistically significant under both assignment strategies at a confidence level higher than .001 ($t_{19} = 49.739$ and $t_{19} = 67.76$ for PCE and PAE, respectively). The expected increase in MPP is confirmed by these results.

Regarding the relative efficiency of the selection methods under each test battery, hypothesis V(b) predicted that greater index efficiency would be detected with 5 tests than with 10 tests. This was assessed first by a comparison of the simple main effects of factor A (at four levels) for the two levels of factor B. Referring to Table 10, it is evident that the main effect of factor A, holding C constant, is larger with 5 tests than with 10 tests in both assignment cases. This suggests that the contribution of test selection method variance is greater under 5 tests than under 10 tests. If this is the case, then the smaller battery allows greater differentiation between methods.

This is overshadowed, however, by the relatively more important finding that differences between methods are significant for both sets of batteries. If the larger battery is assumed to be a less sensitive "test-bed" for exhibiting methodological differences, then the significant finding takes on additional importance. It could even be speculated that the addition of more Project A experimental tests to the battery, beyond the first five, was strengthening the general cognitive ability component since most of the experimental predictors were designed to maximize predictive validity. The display of methodological differences, despite the weak test-bed, enhances the significant results reported for hypothesis II.

Table 10. Simple Main Effects of the Selection Methods by Battery Size

Source of variation	df	MS	F	p
PCE assignment				
A at B_1	3,57	0.0241	106.11	<.001
A at B_2	3,57	0.0049	28.56	<.001
PAE assignment				
A at B_1	3,57	0.0295	117.46	<.001
A at B_2	3,57	0.0058	36.66	<.001

7. Results of Hypothesis VI: Number of Jobs and MPP

Brogden's (1959) model of allocation predicted an increase in the allocation average as the number of jobs increased. Evidence of such an increase was observed here in both assignment situations. In "PCE assignment," an average increase of 10% occurred from job sample 9A to 18 jobs and 6% from job sample 9B to 18 jobs. A Scheffé comparison of the three job sample means across all conditions showed that all the means were significantly different from one another at $p = .05$. More refined t-tests of differences between each mean were all significant at $p = .01$ (see Table 11).

In "PAE assignment," the observed increase averaged 7% from job sample 9A to 18 jobs and 5% from job sample 9B to 18 jobs. A similar comparison as before across the three job samples also showed significant differences between both the 9 job and the 18 job categories at $p = .01$ (see Table 11).

Table 11. t-Tests Between Job Samples 9A, 9B, and 18

Comparison	df	t	p
PCE assignment			
9A v 9B	38	3.783	0.001
9A v 18	38	11.902	<.001
9B v 18	38	8.119	<.001
PAE assignment			
9A v 9B	38	4.466	<.001
9A v 18	38	9.263	<.001
9B v 18	38	4.797	<.001

As expected under hypothesis VI(b), the percentage increase fell below that of 26% predicted by Brogden (1959, Table 2, p. 190)[12] when the number of jobs was doubled from 7 to 14. Several factors explain this difference. Firstly, in Table 2 Brogden's model did not consider the effects of selection, which truncates the applicant population and restricts the range of the assigned population's ability. The potential gain from an increased pool of predicted performance scores from which to assign, therefore, is lessened.

Secondly, Brogden assumes that the dimensionality of the joint predictor-criterion space is equal to the number of jobs (or job families) represented by test composites. It is doubtful that the dimensionality of this space is doubled when going from 9 to 18 jobs using the Project A data. Brogden's model does provide for an increase in PCE when either the multiple correlation coefficient (**R**) is increased and/or the LSE intercorrelations (**r**) are decreased. One would not expect to modify either **R** or **r**, other than through sampling error, when jobs are doubled in number, without adding job families, as in this study. An increase in the number of job families through the shredding of existing families, or by reconstituting families in accordance with differential assignment theory, would increase **R** and decrease **r**, and thus provide a greater gain in PCE than expected here.

8. Results of Hypothesis VII: Criterion Measure, MPP and Index Efficiency

The expected benefit from using the criteria containing hands-on performance measures over criteria without hands-on measures was tested by a comparison of the two 9-job samples. Given that the null hypothesis of no difference between the two 9-job samples means was not directly assessed in the overall F test, a further ANOVA was performed including only levels 1 and 2 of factor C with factors A and B. The main effect of factor C was significant at p = .01 ($F_{1,38}$) =7 3.96) and a t-test of the two sets of means was significant at p = .01 (t_{38} = 3.783 and t_{38} = 4.466 for PCE and PAE assignment, respectively). Although it could be concluded that these two 9-job samples, which differed only in the type of criterion they employed, were significantly different, there was no

---

12 This table provides expected criterion standard scores resulting from making optimal assignments to sets of jobs ranging from 2 to 15, when assignments are made on the basis of full knowledge of these scores. These table entries are readily converted to MPP standard scores for numbers of jobs falling within this range. The effects of doubling 7 jobs can be directly determined from Brogden's table when his assumptions are met, but the effect of doubling 9 jobs can only be inferred by extrapolation. Obviously, doubling 9 jobs would have more effect than doubling 7 jobs.

evidence that this difference was in favor of the 9A sample. The MPP scores for 9A were not consistently higher than the MPP scores for 9B in either assignment case.

More importantly, however, the 9A sample provided a greater distinction between the MPP provided by tests selected by CE as contrasted to SE indices. Such results show that the dimensionality of the joint predictor-criterion space is greater when hands-on criteria are used instead of only job knowledge tests. Referring back to Tables 8 and 9 in this chapter, the F ratio and corresponding t value were larger for the simple effects of A at C_1 (job sample 9A) than at C_2 (job sample 9B) in the "PCE assignment" (Table 8). Furthermore, at least one of the conditions involving the 9B jobs in the "PAE assignment" (Table 9) could detect no significant effects of the indices. It would seem that the relative efficiency of the different test selection methods was better reflected when hands-on criterion measures were employed. However, at least some of the greater sensitivity provided by the 9A data, as compared to the 9B data, is due to the generally larger Ns for this set of jobs. The two average Ns are 448.8 and 326.3, respectively. More importantly, there are three validation samples in 9B smaller than the smallest in 9A.

This result supports the conjecture that the performance of a selection index designed to measure differential validity should be better under a criterion designed to measure more technical job components. This point is emphasized in the study because of the early decision to use the single core technical proficiency (CTP) criterion rather than the composite criterion. The nature of the other MOS-specific CTP measures is such that they would be expected to be complemented and therefore enhanced by the effect of an added job criterion component composed of hands-on measurements of performance.

The increased sensitivity of the 9A job sample raises the possibility that a more accurate test of the performance of PCE and PAE indices could be carried out on a job sample selected to better differentiate across jobs and predicted performance scores. The 19 Project A MOS were selected as broadly representative of the Army's 258 MOS, with particular focus on the combat arms. The state of the art for development of classification efficient measures has not yet progressed to where the two g measures (cognitive g and general adjustment g) can be effectively supplemented in the prediction of performance in the combat arms. Thus, a set of jobs with greater representation of technical specialties would have provided a more efficient test-bed for evaluating the effectiveness of alternative test selection methods for increasing the classification efficiency of the battery.

C. ESTIMATING THE PRACTICAL SIGNIFICANCE OF CHANGES IN MPP

The issue of practical significance questions whether the observed gains in MPP were sufficient to be of practical importance for the use of CE methods of test selection in relevant situations. A translation of the increments in MPP into an estimate of real value provides a better indication of the potential for improving the CE of the operational system. Nord and Schmitz' (1989) simulation of assignment using the Army's Enlisted Personnel Assignment System (EPAS) evaluated the economic benefits of improved MPP across alternative assignment strategies. Their estimates of the costs and benefits of performance in dollar terms provide a basis for extrapolating utility estimates from the results of this study.

In the context of Nord and Schmitz' simulation, a full least squares (FLS) optimal assignment, where individuals were selected and classified to maximize their predicted performance and no quality constraints were imposed, provided an improvement of 0.12 MPP standard scores over the EPAS system, where constraints were enforced and the objective function was to maximize the AA score in the assigned job. The gross economic value of this gain, without considering costs (e.g., recruiting, training, attrition), was estimated to be approximately $200 million for one year (Table 3.19, p. 47).

In the current study, the performance gain from using PCE-oriented test selection methods as opposed to the PSE-oriented method, averaged 0.04 standard deviations over all comparisons, reaching as far as 0.07 and 0.08 in the 5-test conditions. A dollar value of $100 million before costs--half that estimated for a gain of 0.12 MPP standard scores by Nord and Schmitz--describes the range of improvement in PCE suggested by utilization of either PDI or H_d instead of Max-PSE to select 4 or 5 tests for inclusion in a future classification battery.[13] This degree of improvement from a relatively cost-free modification of assignment variables provides a viable alternative to increasing selection standards.

---

[13] It should be noted that Max-PSE is a psychometrically superior test selection index as compared to indices utilized in recent test development studies in which the objective was the improvement of a classification battery. Max-PSE is a difficult competitor to beat. From this point of view the estimated gain obtainable from the use of either PDI or H_d is conservative. However, this estimate is also based on the assumption that the operational selection and classification system used to implement the newly selected test battery will include optimal assignment to jobs, over the past decade this would have been a difficult assumption to fulfill. The implementation of such an ideal system may become more feasible as the Army becomes smaller--a possible consequence of the cold war's end.

IV. OPERATIONAL IMPLICATIONS

A. IMPLEMENTING CLASSIFICATION EFFICIENT TEST SELECTION

The observed gains in MPP from the use of classification efficient test selection methods translate into millions of dollars of increased productivity. The findings of this study provide strong support for a proposed operational change to the present Army system discussed by Zeidner and Johnson (1989b, Chapters 5 and 6), namely, supplementing the existing ASVAB tests with new predictors selected from Project A experimental measures by a test selection index which maximizes PCE rather than predictive validity.

In a previous report, Johnson and Zeidner (1990) reviewed the psychometric principles and the somewhat limited empirical evidence justifying classification efficient methodologies. The present study provides further empirical proof for several aspects of the theoretical foundations of classification efficiency. It does so by way of a rigorously designed simulation of selection and assignment and with the benefit of a systematically derived and unbiased estimate of parameters for the youth population.

The PCE of an operational battery can be improved by increasing the classification efficiency of the tests in the experimental test pool and/or by augmenting the test battery with a set of tests chosen to maximize classification efficiency. The former approach implies a costly development of new tests whose content, unlike the Project A experimental test pool, was preselected or designed to maximize classification efficiency. This task is beyond the research described here. An entire reconstitution of the ASVAB, while not ruled out by the results of this study, is highly unlikely to occur as the result of research bearing on a single service. It is much more likely that the other services would agree to the deletion of one or two tests showing poor classification efficiency, and their replacement by one or two tests with high classification efficiency. It is even more likely that the Army would be permitted to use a two-tiered system in which a supplemental battery for classification purposes is used. The significant gains in MPP under PDI or H_d in the ten-test cases, where differentiation between methods seemed least likely, simply magnifies the certainty that the superiority of the PCE-oriented methods is real.

For the Army, the most promising area for immediate operational change is the augmentation of the 10 ASVAB tests with one or two tests, most likely from the Project A test pool, which maximize PCE. This would require the application of a selection method such as H_d or PDI as described in this experiment. The present results indicate that either would provide a reliable estimate of a test's classification efficiency although we would recommend the use of PDI unless classification efficient factor measures are being sought.

The strong effects observed under the five-test conditions, with gains of 22 percent using hands-on measures of performance, provide great promise for the addition of one or more classification efficient tests to the existing battery. Such a change in the Army's current aptitude areas could be expected to provide over 100 million dollars gain in productivity. These figures are of course based on a comparison with a test selection strategy using the index Max-PSE which, by maximizing the predictive validity of the selected tests, provides a test battery which is much more efficient than the current system of AA composites of ASVAB tests selected by a combination of political considerations and empirical approaches that focused on one composite at a time. Although this study does not pretend to provide a simulation of the present Army system, strong evidence is provided for maximizing a test battery's classification efficiency rather than its predictive validity when the battery is to be used for personnel classification. We would expect a gain of about 40 percent in MPP with the use of FLS composites comprised of classification efficient tests over the present ASVAB FLS composites.

B. INCREASING THE NUMBER OF FLS COMPOSITES AND JOB FAMILIES

Zeidner and Johnson (1989b, Chapter 5) predicted that the PCE of a fixed operational battery can be improved by increasing the number of job families with associated test composites. The present study demonstrates gains of six to ten percent in MPP when the number of jobs is doubled from nine to eighteen. A maximally efficient increase in the number of Army job families and associated predictor composites would be expected to result in between 20 and 40 AAs and associated job families.[14] However, the

[14] Psychometric theory supports the use of as many job families as can be supported by the available validity data when maximization of PCE is the objective. We believe that up to 40 MOS clusters (job families) can be supported by adequate validity information on the existing ASVAB tests ("subtests"). Whether the gain in PCE provided by 40 job families, as compared to 20, is sufficiently large to justify the administration problems inherent in having the additional 20 job families should be determined by a model sampling experiment.

results imply that an initial increase from the current 9 to between 15 and 18 should provide much of this possible increase while being less disruptive to the existing operational system. A two-tiered system of AAs, in which a smaller set is used for counseling and administrative purposes, while a larger set is used within the computer for initial assignment, should be considered in conjunction with a larger, maximally efficient number of AAs.

The findings favoring H_d or PDI over Max-PSE obtained in the sets of nine jobs did not extend to the conditions based on sets of eighteen jobs. While MPP is shown to be greater for 18 jobs than for 9 jobs, the superiority of H_d or PDI over Max-PSE has been blurred in the larger set of jobs. This blurring can be understood as a rational consequence of increasing the importance of the common factor space by doubling the number of variables somewhat evenly across the joint predictor-criterion space without more than a trivial increase in the size of this space (i.e., without increasing the dimensionality). It is understandable that MPP would be increased by the use of more assignment variables (AAs) even though these additional variables represent distinct vectors within the same space. However, it becomes less important that the predictor selected to define these AAs be identified by H_d or PDI instead of Max-PSE.

The results of this study provide preliminary support for the initial phase of research on alternative sets of job families beginning with graduated shredding of the job families (Zeidner and Johnson, 1989b). A further study, currently being conducted, will investigate the effect of creating a greater number of job families featuring a classification efficient job clustering method. By way of a simulation procedure similar to that of this study, it is anticipated that they will demonstrate conclusively performance gains resulting from a job family structure with inherently more classification efficiency.

C. CAPITALIZING ON HIERARCHICAL CLASSIFICATION EFFICIENCY

The Army's current assignment strategy is deficient both in allocation efficiency (AE) and hierarchical classification efficiency (HCE). The unit-weighted, three-test AA composites lack the differential validity which could be provided by the use of FLS composites, and hence limit the AE of the overall process. The present study provides the FLS composites, the most efficient estimates of performance, as the assignment variables, but provides two separate sets of assignment conditions, one with and one without the capacity of providing HCE in the assignment process, focusing on the effects of increasing

HCE. The extent of the gains achieved from capitalizing on HCE (by permitting the FLS composites to remain weighted by their validities) was examined against the effect of a pure allocation efficient process with equal means and standard deviations (and hence no weighting of assignment variables by validities) across jobs.

An overall gain of two percent resulted from the assignment strategy permitting capitalizing on HCE ("PCE assignment") over classification in the absence of HCE (allocation or "PAE assignment"). As in the four-variable model used to demonstrate this effect in Johnson and Zeidner (1990, Appendix 1B), there is some evidence that the combined effects of HC and allocation efficiency increase overall performance. The extent of this difference, however, is no greater than would be achieved by the use of a classification efficient index such as PDI or H_d in the pure allocation case. The benefits of this marginal increase in MPP must be considered against the assumed cost involved in transforming an existing assignment system to one which capitalizes on HC.[15] Hence, on the basis of the present results, modifications to the Army's assignment strategy to encompass HC cannot be given a high priority. In practical terms, more significant gains can be realized from either the use of a PCE-based method of test selection or the use of more AAs.

D. LIMITATIONS AND FUTURE RESEARCH

One crucial factor which may have influenced the results of this research is the nature of the data on which it was based. The Project A data base in general provided an unequalled opportunity to empirically examine issues and theories related to classification which previously would not have been possible. The ideas of Horst and Brogden had remained mostly theoretical and the study by Harris (1967) was conducted using fewer experimental variables than was made possible by the Project A data.

Given the primary hypothesis of this study, however, the emphasis on predictive validity during the development of the experimental Project A predictors may have inhibited the precise effect being analyzed. This suggests that a more sound examination of

---

15 An individual in a key position for influencing the future direction of ASVAB development commented as follows:

> Capitalization on HCE implies layering assignments (assigning all of the most able applicants to the job with the greatest combination of importance and test validity, all of the next most able applicants to the job with the next greatest importance/validity, etc.). Extreme homogeneity among accessions to each job might actually be detrimental as suggested above.

classification efficient test selection methodology would be based on a pool of tests which themselves were constructed to maximize differential validity. In effect, a classification efficient developmental approach should begin with an effort to assemble an experimental battery containing measures that can be expected to expand the joint predictor-criterion space.

Similarly, the selection of the 19 Project A MOS was directed towards a general representation of "key" Army MOS. The inclusion of several MOS (e.g., the combat jobs) that were known to contain more cognitive g and a common "general adjustment to the Army" factor than most other Army jobs reflected the importance of these jobs to the Army. A job sample with inherently greater differentiation between job families would have provided a better sample for detecting differences between classification and selection efficient methods. Zeidner and Johnson (1989b, Chapter 5) discuss a study based on this particular aspect of classification research where alternative degrees of classification efficient job clustering are examined as part of a strategy for improving PCE. This study will examine the loss in utility when the Army job families are reduced to six and the gain if they are increased to twelve for one data set and, for another data set, will compare the amount of PCE provided by job family sets with 9, 16, and 23 members. This should provide a more refined assessment of the effect of increasing FLS composites and their associated job families on classification efficiency.

The ability to generalize the findings of this research to the Army's populations of jobs and incumbents should be viewed within the limitations posed by the methodological nature of the design. Conditions were not constrained to Army-specific requirements (e.g., MOS priorities or job-specific quality goals). Clearly, widespread changes to the current system would require further research designed within the specific requirements of each Service. The advantage of this approach for the purposes of this study is that it allows the superiority of competing methodologies to be isolated from possible policy influences. Realistic quotas imposed on the simulation would have increased the variance of MPP across jobs with the combat jobs being lowered the most. A simulation of all assignment policies would have necessarily included quality distribution goals across jobs. Nord and White (1988) note that one of the effects of constraining an allocation system to meet quality distribution goals is to lower average predicted performance and make it less variable across jobs. Thus, if policies had been more closely simulated, differences between alternative assignment situations would have become harder to detect, and the methodological intent of this research hindered.

Generalizing to jobs in the other military services would seem feasible given that the assignment policies and ASVAB validities for these jobs are comparable to those of the Army.

V. RELATIONSHIP BETWEEN THE RESEARCH FINDINGS AND DIFFERENTIAL ASSIGNMENT THEORY

A. CONCEPTS AND PRINCIPLES OF DIFFERENTIAL ASSIGNMENT THEORY

A major aspect of the importance of these findings is their contribution to the understanding of differential assignment theory (DAT). Also, the findings of this study are best understood in the context of DAT. One purpose of this chapter is to describe how these findings may be used to refine the principles of this theory. DAT and alternative theories relating to selection, placement and classification of personnel are first summarized and the relevance of study results to these theories is examined.

DAT provides a basis for generating a large number of principles applicable to the improvement of operational personnel systems. These principles are obtained as a result of focusing on the gains obtainable from a deliberate and methodologically correct attempt to capitalize on the differing requirements of jobs, using optimal selection and assignment algorithms in an appropriate context. This context includes: (1) appropriate test batteries; (2) best weighted selection and assignment variables; and (3) well structured job families. The psychometric principles of DAT are factual within the constraints of the assumptions necessary to derive them.

The practicality of applying some DAT principles depends on the existence of empirical relationships in a multidimensional joint predictor-criterion (JP-C) space that are non-trivial after sampling error is taken into account, and robust simplifying assumptions required to estimate the solutions of multidimensional integrals. Both conditions can be readily verified or rejected by means of model sampling experiments.

DAT is enriched by broadly based validity generalization (VG) concepts and findings. However, the VG emphasis on a general cognitive ability factor, g, or on g plus one or two additional group factors, prevalent among strong proponents of VG theory, is not a requisite characteristic of DAT. While the theory is not restricted to any particular factor structure, the assumption of a non-trivial degree of multidimensionality in the JP-C space is essential.

DAT principles originate from either mathematical derivations or empirical observations; they have been identified principally by attending to four organizing concepts. These concepts involve making assumptions regarding the relationships that can be found among predictors, criterion variables, and dollar-based benefits in the context of well designed personnel utilization systems. Assumptions inherent in these concepts permit us to speak only of a theory of differential assignment rather than a simple set of algebraic derivations and/or empirical data supporting the principles.

The first of these concepts holds that there is a set of quantitative principles governing: (1) selection of predictors for inclusion in experimental pools of predictor measures, in operational batteries, or in sets of test composites, to maximize selection, placement, and/or classification benefits; (2) the optimal constitution of job families; and (3) the selection of strategies (e.g., one-stage versus two-stage) and algorithms for the selection, assignment, or simultaneous selection-assignment of personnel. These principles must be followed in order to maximize benefits, and must be considered along with costs to maximize system utility.

The second concept is based on the assumption that utility models, balancing costs and benefits, provide the best approach for specifying personnel selection and classification policies and procedures for operational systems. It is further assumed that benefits attainable from such personnel systems are most appropriately measured in terms of predicted performance, possibly weighted by the dollar value of a given level of performance for each job.

The third concept incorporates the well-known principle that the benefits of both selection and classification procedures can under certain circumstances be maximized by the use of the same test composites for selection and/or classification; i.e., the "best" test composites corresponding to the same job or job family having the same tests and weights. The hard decision as to whether selection or classification efficiency should be maximized can be avoided, but only if each composite is a full least square (FLS) estimate based on the "full" battery (i.e., containing all the available tests).

An assignment system that uses FLS composites for all selection and assignment variables and optimally assigns personnel separately to all jobs, is an optimal assignment system for both selection and classification. When a subset of the experimental test pool is selected for inclusion in the operational battery, or when simplified composites that are not FLS composites are selected from the operational battery for utilization in selection or

assignment variables, or when jobs are clustered into a small number of job families, then less efficient substitutions have been made into this potentially optimal assignment system. In making these substitutions, it is inevitable that classification efficiency will suffer if selection efficiency, relying on the maximization of predictive validity, is optimized without specific consideration of classification efficiency.

This third concept includes the assumption that non-trivial increases in benefits, or mean predicted performance (MPP) can result from deliberate attempts to maximize classification efficiency. Findings of this study show that maximizing predictive validity, without consideration of the effects of classification efficiency, will under some circumstances, prevent maximizing MPP; a finding consistent with other evidence. This concept contrasts to a point of view, one particularly prevalent among VG proponents, holding that the appearance of classification efficiency in the absence of hierarchical classification (HC) effects is almost always due to sampling error or to the effects of various biases.

The fourth concept holds that the computer age makes it practical to implement operationally any multidimensional selection and assignment strategy and algorithm that can be shown to provide gains in MPP, regardless of the complexity of the process.

DAT provides the basis for generating a large number of principles. (See Johnson and Zeidner (1990) for a comprehensive treatment of these principles.) Until recently, substantial evidence that real world relationships predicted by these principles would produce non-trivial effects (in the context of realistic measures of benefits and controlling for sampling error) has for the most part been lacking. For practical reasons, such evidence could be produced only through simulation experiments such as the one used in this study. Ten representative DAT principles are summarized below.

1. Use of FLS Composites

Probably the most important of these DAT principles is that the "best" selection and assignment variable, (i.e., "best" for maximizing either selection or classification efficiency), is an FLS composite. When for practical reasons predictors are selected from an experimental pool for inclusion in an operational test battery, and/or further selection of tests is made to provide smaller test composites (e.g., as in aptitude area test composites of the services), different sets of tests will maximize PSE and PCE. Also, as jobs are clustered into job families, differe· ıpproaches are required to provide families that

optimize PCE as contrasted to PSE. Whether the gain in PCE and the loss in PSE resulting from the use of classification efficient test selection methods is substantial, moderate or negligible is partially determined in the present study.

2. Measurement of Potential Classification Efficiency (PCE)

This principle states how PCE can be measured. Benefits obtained from a classification process can be determined only after assignment of individuals or simulation entities have been accomplished using the specified assignment strategy. Gains or losses in PCE cannot be determined solely as a function of predictive validity or by such techniques as path analysis. For this reason, most empirical findings relating to the effect on PCE of various conditions (e.g., method of test selection, number of tests in composites, number of assignment variables) have been obtained by means of a personnel system simulation that includes assignment of personnel, and also usually includes a selection procedure.

3. Estimating Mean Predicted Performance (MPP) as a Function of R, r, and f(m)

The potential gain in PCE from the application of several DAT principles can be stated in terms of Brogden's formulation of MPP (Brogden, 1959). Brogden's model and the extension of this model by Johnson and Zeidner (1990) provide approximate roles for: (1) average predictive validity, **R**, (2) the average intercorrelation among FLS composites, **r**, and (3) an order function, f(**m**), reflecting the effect of the number of jobs, or job families (**m**), and corresponding test composites. The MPP standard score is predicted by Brogden's formula: MPP = f(**m**) times **R** times $(1\text{-}\mathbf{r})^{1/2}$. It is notable that **r** can be, and usually is, quite large without causing the magnitude of PCE to become negligible.

4. Source of PCE

PCE is derived from two often competing sources: allocation effects due to differential validity and hierarchical classification (HC) effects due to the matching of hierarchical layers of predictor scores to the mean benefits attached to each job. HC effects can be present even if there is only a single predictor composite used for all jobs, if hierarchical layering effects of mean validities or job values match the layers of benefits attached to jobs. It should be noted that the same linear programming (LP) assignment algorithm may be used whether PCE consists only of allocation effects, HC effects, or a combination.

Allocation efficiency results from the variance within each individual of predicted performance for different jobs. The presence of such variance is demonstrated by differential validity, higher validity for each composite with its associated job criterion than for the other criterion measures. The potential end result of making optimal assignments using a set of composites possessing differential validity is analogous to improving the SR for those selected for assignment to each job (or job family). That is, when the pool to be assigned is separately rank ordered on predicted performance (PP) for each job family, only the upper portion of the continuum is assigned to that job. This analogy does not hold for HC effects, since with only one predictor (or a single dimension in the test battery) the assignees are deposited in successive layers with no improvement in the average percent of assignees ending up in the upper portion of the PP continuum for each job family. Although HC effects provide no improvement along lines analogous to SR, either HC or allocation effects in the assignment process can provide an increase in MPP.

5. The Dimensionality of the Joint Predictor-Criterion (JP-C) Space

Factor analysis is a significant tool in classification systems research and has an important role in DAT. The JP-C space, rather than test space or common factor space, a subset of test space, is the domain of interest for investigating the utility of selection and classification processes. It is in this joint space that H_d and H_a is measured and that factors for maximizing either selection or classification effects can be identified.

6. The Number of Predictors in a Composite

The incremental gains provided by adding more tests, either to each composite or to the operational battery, approach zero much faster for unidimensional than for multidimensional selection and/or classification processes. Thus, the best five tests selected from an experimental pool of predictors and used in an LSE of the criterion can be expected to approach the maximum validity obtainable for that predictor pool as computed in independent samples (i.e., for the computing of "cross validities"). Conversely, an unbiased measure of PCE will continue to increase as the size of the operational battery is increased to several times this number (assuming the tests are appropriately selected and there are a half dozen or more fairly heterogeneous job families).

7. Clustering Jobs into Families

In general, increasing the number of assignment composites and associated job families adds to PCE. However, there are a number of different ways in which the number of assignment variables can be increased. The existing operational job families could be shredded into relatively homogeneous smaller job clusters; the total set of jobs might be reconstituted into classification efficient clusters; or the number of job families might be increased without expanding the JP-C space. The magnitude of a gain in PCE resulting from an increase in the number of assignment composites and job families will depend on the method used to provide more job families, the heterogeneity of jobs in the JP-C space, and the number of composites from which an increase is being accomplished.

Brogden provides an estimate of the potential gain from expanding the JP-C space by adding additional jobs or job families with associated FLS composites possessing an independent component. This model predicts that an increase of the number of FLS composites from 5 to 10 will provide a percentage gain in MPP of 33 percent. This magnitude of gain is more than would be predicted from the increase of the number of jobs from 9 to 18, as is the case in the present study, where the additional jobs are in the same JP-C space. However, it is less than would be expected for such an increase in job families using a method that increases the average validity and decreases the average intercorrelations of the FLS composites.

To compare the circumstances under which the assignment variables in this study were doubled in number with the circumstances implied by Brogden's model, consider a set of four jobs from each of four existing Army job families (i.e., families corresponding to Army aptitude area composites). If FLS composites for each of these jobs are used as assignment variables and then augmented by four more FLS composites associated with four additional Army jobs, each associated with a different job family than the first four jobs, we would expect the gain in MPP (PCE) to approximate that predicted by Brogden's model. However, if the number of jobs and assignment variables were to be doubled in a different manner, by selecting an additional job from each of the first four job families, the circumstances would closely resemble those of this study where the number of jobs was doubled (i.e., from 9 to 18). We would expect a much smaller gain under these latter circumstances because the added jobs do not appreciably expand the scope of the JP-C space. However, we do expect a non-trivial increase because of the increased number of job continuums on which PP scores can be rank ordered and only the best accepted for assignment in a manner analogous to obtaining an improved overall selection ratio.

8. Optimal Assignment of Individuals to Jobs

There are a number of crucial DAT principles bearing on the effect that constraints have on overall MPP standard scores and on the MPP standard scores of those assigned to each job (i.e., quality distribution). For example, the optimal assignment of individuals, to meet job quotas and maximize total MPP, can be obtained by assigning each individual to his highest adjusted predicted performance (PP) score; adjusted PP scores are defined as PP scores plus appropriate job constants (i.e. dual parameters). A number of variables affect the relative magnitude of these dual parameters. Other things being equal, the jobs with the larger quotas have the larger dual parameter. Similarly, the jobs whose PP variables have the highest average correlations with other PP scores have larger dual parameters.

9. Selection and/or Classification Strategies

There are many alternative selection and/or classification strategies including: (1) selecting applicants into the organization on a single predictor and then classifying those not rejected, using multiple assignment variables, to specific jobs (the traditional two-stage strategy); (2) using one predictor at a time in a sequence of select - train - select - train - select - place (a multiple hurdle strategy as used in the Army's helicopter pilot program); and (3) simultaneous selection and optimal classification to multiple jobs, as in the single stage MDS algorithm described by Johnson and Zeidner (1990).

The MDS algorithm is best understood in the context of Brogden's model (1959) where each predictor is an FLS composite yielding a score that divides into a general and a unique component. The increased MPP resulting from classifying a selected set of applicants, instead of classifying all applicants (with no rejections), is in Brogden's model the result of an assignment strategy in which applicants are simultaneously selected and classified to jobs using only the unique components, **u**, of the FLS composites. Removing the g component produces no adverse effect on the MPP resulting from classification, but obviously loses the gain in MPP that would result from also selecting on **g**, instead of only on **u**. Johnson and Zeidner (1990) modify Brogden's model to reflect a strategy in which selection is on the g component and classification remains (in a second stage) on the **u** component of each FLS composite score. This latter strategy increase the gain in MPP obtained from selection effects when g constitutes a large part of each score (as is usually the case). A simultaneous, one stage, selection-classification strategy (i.e., MDS) such as

is implied by Brogden's model--but using the complete FLS composite score instead of just **u**--provides the maximum possible amount of MPP.

DAT points to MDS as the optimal selection-classification strategy; unfortunately MDS has never been implemented operationally, possibly because DAT is poorly understood and possibly because the power of modern computer technology has been only partially applied to operational selection and assignment systems.

10. The Use of Factor Scores as Assignment Variables

Factor scores provide a means of maximizing the PCE obtainable from a small set of **k** composite scores for a fixed structure of **m** jobs or job families (**k** < **m**) to be recorded in a soldier's personnel record. Given that this small number (**k**) of recorded assignment variables can be computed from a larger number of predictors in an operational test battery, and each assignment variable is an LSE (based on the **k** recorded composites) predicting one of the **m** job criteria, the PCE attainable from recorded test composites is smaller than the PCE attain from use of the full operational battery to compute each assignment variable. This decrement in PCE may not be large enough to outweigh the advantage or recording and using the scores of the **k** composites noted above. These **k** composites would be factor scores, that is, FLS estimates of factors obtained so as to maximize classification effectiveness. Such a factor solution will be described below.

A factor solution, providing factor coefficients for the predictors against factors for which the factor contributions in the JP-C space are also each factor's contribution to H_d, is first computed in the manner described by Johnson and Zeidner (1990). H_d can be maximized for a given number of factors by retaining those with the largest contributions. Factor scores computed as FLS estimates of these retained factors maximize PCE to the extent that H_d approximates PCE.

Classification efficient factors are obtained by rotating a PC solution obtained in the JP-C space to a solution for which each factor successively maximizes H_d. The computation of predictor weights that provide an FLS estimate of these classification efficient factors readily follows. The use of factor scores in an hypothetical two-tiered operational system is described in Zeidner and Johnson (1989b).

Having reviewed organizing concepts and representative principles of DAT, DAT can be summarized as a collection of concepts and principles concerned with how to measure and optimize selection, placement, and classification systems. This theory differs

from alternative theories with the same objective in its emphasis on utility and methods for capitalizing on the multidimensionality of the joint predictor-criterion space. Evolving from the acceptance of MPP as the basis for measuring the benefits component of utility, DAT uses MPP as the means of evaluating all selection, placement, and classification approaches. DAT most obviously differs from any theory which considers the benefits accruing from classification systems to be the sole function of predictive validity. However, DAT equally differs from any theory which does not recognize the validity and importance of the validity generalization movement in its contradiction of "situational specificity theory."

B. A COMPARISON OF DIFFERENTIAL ASSIGNMENT THEORY WITH ALTERNATIVE THEORIES

There are a number of alternative theories to DAT, including, at one extreme, one aptly described by Ghiselli (1959) as the "situational specific theory," and, at the other, "general factor" theory. General factor theory ascribes an all pervasive, overriding dominance of a single general cognitive ability measure in the JP-C space, and concedes no reliable importance to any other measure in that joint space. General factor theory is referred to as general cognitive aptitude theory by Schmidt, Hunter, and Larson (1988).

Those holding to the latter theory can be counted upon to support VG concepts, but VG proponents do not necessarily support the "general factor theory" or general cognitive aptitude theory as described by Schmidt et al. (1988); thus, these two schools of thought are by no means identical. DAT also incorporates the empirical findings and lessons learned from VG theorists concerning: (1) the strength of validity generalization across jobs; (2) the desirability of using adequate sized samples to estimate validities (often obtained by referencing broader universes); and (3) the difficulties to be expected in the development of operational systems having the desired level of classification efficiency.

Situational specificity proponents mandate that empirical validation is required in each situation where there is a difference resulting from the characteristics of jobs, organizations or criteria. This requirement must be met to justify the use of a predetermined predictor measure as a selection, placement, or classification device. Situational specificity theorists may argue a "scientist" would not presume to use the results of a previous validation study for a slightly different job, criterion, location, or applicant group for which all situational variables were not exact fits. Thus it would not be possible to develop test composites for use in ongoing selection and classification situations for fear

that situational changes would have occurred since the last validation, validities could not be presumed to generalize to closely related jobs within the same job families, and general principles could not be drawn from personnel measurement research.

If situational specificity views were true, that is, if no more than a trivial amount of validity generalization exists, the potential for operational classification systems would be great indeed. Unfortunately, this potential, like selection potential, would be impossible to realize without an impractical number of situational specific validation studies. If in rejecting the situational specific point of view, one swings entirely to the unidimensional view of JP-C space, the feasibility of implementing systems having selection effectiveness is very good, but the feasibility of implementing an effective allocation system disappears. That is, the implementation of any operational system possessing an acceptable level of classification efficiency, other than that due to HC effects, becomes highly unlikely because of the expense involved in conducting the required number of situational specific validation studies.

DAT is compatible with the prevailing evidence that a single principal component factor often explains 60 to 80 percent of the total factor contribution in the JP-C space. The remaining 20 to 40 percent of the total factor contributions can provide the basis for non-trivial classification efficiency in the space bounded by the smaller factors and the largest factor (g). However, DAT does not require the acceptance of a particular factor structure (e.g., a single factor for explaining general cognitive ability) as explanatory of the relationships among cognitive or non-cognitive predictors in the JP-C space. The factor structure in either test space or common factor space, as defined by predictors only, is not relevant to DAT.

Brogden's (1959) model assumes a factor structure in the JP-C space corresponding to Spearman's two factor theory. This assumption avoids the necessity of solving intractable multivariate integrals, thus simplifying the computations required to estimate PCE for varying classification system characteristics. The relationships among Brogden's "predictors" (FLS composites providing PP scores) are entirely explained by a single general factor, but he also assumes as many unique factors as there are jobs. We believe Brogden did not hold that Spearman's two factor theory was an accurate representation of empirical relationships. However, Brogden's model that implies the factor structure of Spearman's "two factor theory" structure with his assumptions (although Brogden does not mention a factor structure) provides a clever approximation of the

benefits that can be reasonably expected to result from a classification system. It is likely that the model is usefully robust when mean **R** and mean **r** are used to enter his tables. The degree of robustness could be readily measured using scores from model sampling in a simulation of personnel selection-classification systems that deviate to varying degrees from Brogden's assumptions. Evidence of a useful amount of robustness is provided here.

Schmidt, Hunter, and Larson (1988) describe three theories entirely in terms of predictive validity. They are obviously proponents of the general cognitive ability theory and, to a lesser extent, of group aptitude theory. Schmidt et al. (1988) provide data that argues against the adequacy of specific aptitude theory.

The critical distinction between results in support of general cognitive ability theory, as contrasted with contradictory results in support of specific aptitude theory, lies in which of the following provides the largest "cross-validity": (1) either AAs or LSEs based on a small number of best predictors for each job; or (2) a predetermined measure of general cognitive ability. They make a similar comparison between the validities of: (1) LSEs based on tests (as dependent variables), and (2) LSEs based on test composites (as dependent variables). They obtain similar findings favoring the general cognitive aptitude theory as compared to the group factor theory. However, it is highly questionable whether group factor scores are adequately represented by the use of such test composites as used in the Schmidt et al. study.

While it could be argued that the use of only predictive validity in defining the three theories in the Schmidt et al. study is irrelevant to DAT, they are possibly the most influential of the alternatives to DAT and have guided many investigators in the development of classification systems. Those interested in increasing the classification efficiency of operational systems should oppose the tenets of all three theories because of their mutual assumption that achieving gains in predictive validity is appropriately the primary goal in the design of classification systems and strategies. We believe that the lack of PAE in the existing ASVAB and its operational composites is largely due to dependence on one or more of these erroneous theories.

While the specific aptitude theory encourages the belief that suitable classification measures can be developed without the use of effective indices for measuring PCE, general cognitive theory on the other hand, encourages the belief that the very attempt to develop classification efficient measures is a hopeless cause. Thus we are equally opposed to both.

Expectations of results from this study and its companion studies differ depending on whether DAT or one of the alternative theories are preferred as the best organizing and explanatory framework. For example, those emphasizing the primacy of **g** in both selection and classification, might expect all gains provided by ten predictors over five to disappear in cross samples (as an artifact of error variance). Similarly, this theory would predict Max-PSE to be as good an index for use in selecting tests for an operational classification battery as either PDI or H_d, and should not predict greater PCE from the use of 18 instead of 9 jobs and corresponding FLS estimates as assignment variables. Our results as indicated, controlling the effects of sampling error, show the expectations of "g theorists" to be inappropriate in the context of the classification process.

C. POTENTIAL CONTRIBUTION OF DIFFERENTIAL ASSIGNMENT THEORY TO OPERATIONAL SELECTION AND ASSIGNMENT SYSTEMS

DAT provides a basis for recommending a number of changes in the overall strategy for selecting and assigning personnel in the military services. Four major changes are listed in the order of their potential contribution to a larger MPP.

1. Improve test composites used for selection and assignment by:
   - (a) using FLS composites for existing job families with each composite converted to AAs having means of 100 and standard deviations of 20; or, alternatively,
   - (b) using unconverted FLS composites permitting capitalization on HC effects.
2. Improve classification efficiency of job families by:
   - (a) increasing the number of job families by shredding the nine job families associated with the Army AAs; or, alternatively,
   - (b) maximizing classification efficiency of job families by increasing and reconstituting families in accordance with DAT.
3. Improve predictor content by:
   - (a) providing preliminary studies for selecting classification efficient predictor measures; and
   - (b) selecting a classification efficient operational battery.
4. Improve selection and assignment strategy by:
   - (a) optimizing the traditional two-stage strategy; or

(b) utilizing simultaneous selection-classification (single stage) strategy with an MDS algorithm; or

(c) utilizing a two-tiered approach in which the first tier uses FLS estimates for a maximum number of job families and the second tier uses a small number of factor scores recorded in the personnel record for job counseling, cut scores, and related purposes.

Simulation results obtained by Sorenson (1965a) and Nord and Schmitz (1989) are consistent with Brogden's proof (Brogden, 1955) that the maximum gain from any one source is provided by using FLS composites. When the Army AA composites used in conjunction with an optimal two-stage strategy and an optimal assignment algorithm are used as the basis of comparison, the percentage gains over chance assignment provided by using FLS composites as assignment variables would lie somewhere between 70 and 100 percent. The further gain provided by capitalizing on HC could range from a trivial amount to 10 or 20 percent. An estimate based on the "four variable model" described in Appendix 1B of Johnson and Zeidner (1990) indicates that a 10 percent gain in PCE could be expected if the Army classification battery was permitted to utilize HC effects. The results reported here suggest the gains obtainable from HC effects in an improved classification battery could range from 10 percent to 20 percent. More direct evidence based on fewer assumptions will be provided in a related GWU study investigating the magnitude of contributions to PCE by HC effects.

DAT principles predict that further gains obtainable from shredding the existing job families into a larger number, over and above those obtainable from the use of FLS composites corresponding to the existing families, would be 20 to 40 percent over chance for an increase of from 9 to 18 jobs. This study shows that doubling the number of job families (each with separate associated assignment composites) by adding another job from each job family, without including all jobs in both sets of 9 and 18 job families, provides a gain in classification efficiency of around 10 percent. We believe that using an optimal reconstitution of a larger number of jobs, with all jobs included in each set of job families, to increase the number of job families from 9 to 18 can be expected to provide a total gain over chance assignment of 30 to 50 percent, an increase that can be added to the gain of from 70 to 100 percent over chance obtainable from substituting FLS composites for AAs while using the existing job families.

The gains obtainable from selecting classification efficient tests from an appropriate pool of experimental predictors, preselected or developed to be classification efficient, are

probably the most difficult to estimate. In the *absence of an appropriate pool of experimental predictors*, algebraic techniques, simulation models, or any number of empirical studies on the extant data cannot provide convincing estimates. As more evidence is accumulated, the possibility that a creative development of new predictors would provide a sizable further gain remains a credible possibility. Methods for detecting new classification efficient predictors within the limitations of current research efforts are provided by Zeidner and Johnson (1989) and Johnson and Zeidner (1990).

The present study employs estimates from a credible universe for use in generating samples of entities through the use of intercorrelations and validates obtained from the Project A experimental pool of predictors. This pool seems to have been assembled with the goal of increasing predictive, rather than differential validity. The experimental variables include biographical and interest inventories, perceptual tasks, and "video game" type tasks. While Project A investigators succeeded in increasing the heterogeneity of the predictor pool as compared to the ASVAB, the manner in which this content heterogeneity was increased would not necessarily result in improved differential validity. Most of the experimental non-cognitive predictors appear equally relevant to many, if not all, Army job families.

There is undoubtedly a non-cognitive **g**, with similar psychometric characteristics as the better known general cognitive ability measure (cognitive **g**). The lack of differential validity, that is having the same level of validity across all jobs in an organization, is characteristic of both cognitive and non-cognitive general measures (cognitive and non-cognitive **g**). Other than through HC effects, neither **g** contributes to classification efficiency.

Johnson, Klieger, and Frankfeldt (1958) and Johnson and Kotula (1958) describe the tendency of self-description tests (including those with interest and biographical items) to equally predict performance in most other jobs, even when scored with empirical keys developed against performance in specific jobs, unless the tests are specifically designed to be differential. The Army Classification Battery self-description test, titled the Army Classification Inventory (CI), provides an excellent example of non-cognitive **g**. This 125 item test was developed using combat criteria and combat potential ratings collected in combat training situations.

Classification efficient tests should have their highest corrected validates in the job family for which the test was developed and an appreciably lower average validity in all

other job families. In order to make this kind of comparison we use data provided by Sorenson (1965, pp. 91-20) to pair tests to their corresponding job families in order to make a thumb nail evaluation of their classification efficiency. For this purpose we pair the CI with the two job families then making up the combat arms (IN and AE); Arithmetic Reasoning (AR), the best measure of cognitive **g** in the JP-C space, is paired with the general technical job family (GT); Automotive Information (AI) is paired with the mechanical repair family (MM); Electronics Information (EI) with the electronics family (EL); and Army Clerical Speed (ACS) with the clerical family (CL). We note that CI has an average validity of .28 in the combat arms and in the other families, an average validity of .46, with a range of between .41 and .59. Similarly, AR has an average validity of .65 in GT and, in the remaining families, an average validity of .54, with a range of between .43 and .69.

Validities are corrected for restriction in range, but not for attenuation. The combat arms criteria are fairly reliable multiple ratings (**r** approximately equals .60), while the other families used school performance measures as criteria. The sample sizes were large, totalling 11,807. Even after making rough corrections for the differing criterion reliabilities, the measure of non-cognitive **g** appears to have generalized at least as much as the measure of cognitive **g**.

We now look at the validities of the three tests for which we expect to obtain differential validity. AI has an average validity of .65 in MM and an average validity of .36 in all other families; EI has an average validity of .60 in EL and an average validity of .43 in all other families; and, ACS has an average validity of .58 in CL and an average validity of .38 in all other families. The difference between the differential validities of the two **g** measures and those of these three job oriented measures is striking.

The non-cognitive **g** found in Project A experimental measures includes general adjustment to the organization (e.g., to the Army) and other job performance characteristics (e.g., achievement motivation) that are equally valuable across all jobs in the organization. While achieving predictor heterogeneity in this manner may improve selection efficiency (predictive validity), such heterogeneity will contribute little to classification efficiency. The results of this present study would have undoubtedly shown a greater difference between the effects of test selection using classification efficient versus selection efficient indices if the Project A test development process placed a greater emphasis on finding classification efficient predictors for the experimental pool.

Our study supports the DAT prediction that the use of PDI or H_d indices rather than Max-PSE to select tests for an operational battery increases PCE. This finding indirectly supports a prediction that attention to similar characteristics in selecting tests for the experimental pool would increase the PCE of the pool. The operational battery might profit more from the consideration of PCE in an earlier stage of research, when predictors are being developed or selected for inclusion in the experimental pool, than from a later selection of tests from a pool that has already been limited to predictors intended for maximizing predictive validity.

The recommended fourth change derived from DAT calls for optimizing the selection and assignment strategy. The traditional distinct procedures of first being selected by AFQT and then being assigned to a job using a set of composite scores from the ASVAB in a later assignment process is far from optimal. Total MPP resulting from both processes is raised, while achieving the same job quotas and selection ratio (SR) by: (1) substituting a FLS measure for AFQT, and (2) using a lower cut score on this selection measure, in conjunction with (3) a higher cut score on the AA associated with each entry job.

Total MPP can be maximized by raising the cut scores on a separate FLS measure for each entry job to the point where the use of either AFQT or the FLS measure governing entry into the organization has a trivial effect. This can be accomplished through use of the MDS algorithm described by Johnson and Zeidner (1990).

This study provides direct support for the implementation of changes 1 and 2b above and indirect support for accomplishing change 2a. The companion studies in progress will provide model sampling evidence pertaining to changes 2, 3, and 4.

D. RESEARCH FINDINGS IN THE CONTEXT OF DIFFERENTIAL ASSIGNMENT THEORY

DAT is appreciably shaped by the results of the present study; we anticipate further refinement in DAT to result from the completion of the ongoing companion studies. The higher MPP provided by both PDI and H_d over Max-PSE and the superiority of ten tests over five tests both provide strong evidence for the credibility of the DAT assumption of a non-trivial degree of multidimensionality in the JP-C space. The higher MPP provided by assignment to 18 jobs, as compared to 9 jobs, supports the accuracy and robustness of Brogden's model (1959).

There are other more subtle conclusions that can be drawn from the present study that were unexpected, and thus more exciting with respect to the increased understanding of DAT. These unanticipated results relate more to the intensity of effects than to their direction. As is generally true, the design of this study has less sensitivity for detecting relationships that have not been hypothesized as important before the collection of data. For example, DAT predicts the advantage of PDI over H_d to not only exist but to be more in evidence when assignment is to 9 jobs than when assignment is to 18 jobs. This reflects the decreasing opportunity for H_d to erroneously capitalize on the interaction of a job and a predictor to select a different test than would be selected by PDI as the number of jobs (or job families) and corresponding composites increases.

Most Army researchers in the 1960s believed that differential validity and predictive validity were reduced when supervisory ratings rather than school grades were used as criteria. In that era, school grades tended to be normative rather than criterion referenced. Today's Army school grades are usually less reliable than formerly, particularly at the upper end of the distribution, and are designed to minimize student failures. Many of the "passing" students later attain unacceptable low scores on the skill qualification test (SQT). We do not quarrel with this training philosophy, but suggest that the school grades of earlier decades were a superior measure for research purposes.

The SQT is now a less reliable measure of performance because of the exclusion of a more difficult type of question from the test. The more difficult items associated with the next higher skill level once, in earlier SQTs, made up 25% of the items of a SQT. It now appears necessary to develop tailored "hands-on" performance tests and achievement tests to obtain the psychometric characteristics of earlier school performance and SQT measures.

Of the five criterion components used by the Project A investigators to measure job performance, four are measures of general soldier performance reflecting characteristics which cut across jobs and thus have little relevance to classification. If these four general criterion components were used without the MOS specific component there could be no evaluation of personnel classification. If these four components were to be designated as the only components of interest to policymakers, one would have to conclude that any further effort to maintain a classification system should be discontinued.

Only one of the five Project A criterion components measures specific performance in a given MOS as contrasted with general military duties. The differential validity is greater against this component than against the sum of all five components. We may

further predict that there will be more differential validity when the MOS specific criterion component includes a hands-on component than when this component consists primarily of ratings and paper and pencil tests.

Assuming the presence of greater differential validity for the set of 9 jobs with criterion variables that includes hands-on criterion measures (as contrasted with the other 9 jobs), we would expect greater sensitivity under these conditions to the use of classification efficient indices foı test selection. In contrast, the use of 18 jobs instead of 9, while providing a larger MPP, also permits a larger contribution of the **g** factor (as discussed earlier), thus decreasing the sensitivity to the advantage of using classification efficient indices. It is logical that a superior index provides a detectable superiority in selecting tests that will provide greater PCE over alternate indices under the more sensitive conditions, and may fail to reject the null hypothesis under the less sensitive conditions.

As noted above, there is reason to believe that the experimental condition involving 18 jobs also provides less sensitivity to the advantage it is hypothesized PDI has over H_d. Furthermore we would predict from DAT principles that a sufficient sample size in a model sampling experiment would provide the detectability of the effect of such conditions as type of criterion and number of jobs with respect to the significance level that the hypothesis of no difference between PDI and H_d is rejected. We did not predict that 20 samples (replications) in each cell (as in the present study) would be large enough to provide this distinction.

With the hindsight of having seen the results, we now hypothesize that: (1) another 20 replications in the cell representing the conditions of 5 tests, hands-on criteria and 9 jobs would again, as in this study, provide statistical significance, and (2) another 20 replications in the cell representing the conditions of 5 tests with 18 jobs, plus a third set of 20 replications in the cell representing 5 tests and 18 jobs, would not provide statistical significance for a difference in favor of H_d over PDI, (separately or together) with the 60 replications we already have in these three cells. We could legitimately test such a compound hypothesis as if we had determined in a back-sample which comparisons and hypotheses we wished to make in a cross-sample. Our own conclusions regarding the relative merits of PDI and H_d will have to await the collection of this further model sampling data. Until then we recommend the use of PDI when there is no special reason to use H_d, but cannot provide evidence for this conclusion in terms of a rejected null hypothesis.

VI. SUMMARY AND CONCLUSIONS

This study is the first in a series designed to examine methods for improving the classification efficiency of the Army's selection and classification system. In the present experiment, the classification utility of alternative methods of test selection was examined in the context of other factors influencing classification efficiency. In general, five major procedures were involved in determining the potential improvement in classification efficiency (CE) under each approach: (1) selection of alternative test batteries from a pool of 29 predictors whose intercorrelations and validity coefficients for 19 jobs were based on a sample of 7045 subjects; (2) generation of synthetic test scores for the selected batteries using a model sampling technique; (3) truncation of each synthetic sample of test scores using a selection ratio of 0.70; (4) optimal assignment of the synthetic sample to MOS based on the test scores; and (5) evaluation of alternative test selection methods under different conditions using the results of optimal assignment (mean predicted performance (MPP) standard scores).

The cross-validation design eliminated the possibility of bias both from back-sample validities and further correlated error between regression weights calculated on the same samples. Evaluation weights were based on the designated universe (the corrected Project A empirical sample) and assignment weights on a sample drawn from this universe. Hence, observed effects could not be attributed to bias or uncontrolled sources of error.

Evaluation of the MPP standard scores supported the primary experimental hypothesis that CE methods of test selection would lead to greater MPP in an assigned group than a selection efficient method. This finding was replicated across assignment strategies with and without a hierarchical classification effect, using five or ten tests in the battery and across nine jobs with and without "hands-on" measures of performance. Reversals in the relative predicted direction of MPP were not statistically significant and, therefore, could not be interpreted as indicating a lack of superiority of PDI, H_d or the combined effect of the two under certain conditions.

Tests of the secondary hypotheses produced mixed results. The advantages of PDI as an alternative to H_d were not observed at the level of significance chosen to interpret the results, although the absolute MPPs under PDI were marginally higher in half the cases and

overall at the .05 level. It is concluded that either index can be used, with confidence, to estimate and improve PCE, although for theoretical reasons we would recommend the use of PDI when a factorial model is not being used.

Improvements in MPP were of a practically significant magnitude as gauged by Nord and Schmitz' (1989) estimates of the dollar value of predicted performance standard scores. This suggested a substantial economic benefit in adopting classification efficient strategies. It was also evident that these economic benefits could be derived at relatively little cost since MPP was significantly improved without the need for an entirely new "maximum CE" battery or an assignment strategy which capitalized on hierarchical classification.

The omission of policy constraints precludes drawing definitive operational recommendations for the Army based on these findings. However, test selection methodology to achieve classification efficiency is recommended. This study has successfully shown that, using either H_d or PDI, PCE can be increased with some economic gain. Additionally, the study has shown that classification efficiency significantly increases when the number of FLS predictors is increased from five to ten; the use of hands-on criteria provides greater differential validity effects across jobs than can be obtained through the use of rating criteria; and doubling the number of Project A jobs from nine to eighteen provides an increase in overall MPP.

GLOSSARY

allocation efficiency--The gain in benefit over random assignment obtained from an optimal assignment process attributable to differential validity.

allocation process--Classification that capitalizes on differential job validity.

attenuation[a] --The reduction of a correlation or regression coefficient from its theoretical true value due to the imperfect reliability of one or both measures entering into the relationship.

battery[a] --A set of tests standardized on the same population, so that norm-referenced scores on the several tests can be compared or u^ed in combination for decision making.

benefit--A theoretically desirable measure of performance that is value-weighted for jobs and validity in terms of an appropriate metric; when the benefit measure is correctly combined with costs, it provides a measure of utility.

classification--The matching of individuals and jobs in an organization with the goal of maximizing aggregate performance; it requires multiple predictors jointly measuring more than one dimension and multidimensional job criteria.

classification battery--A battery of tests used operationally to classify personnel.

classification efficiency--The gain in benefits over random assignment obtained from an actual assignment process in use that is attributable to allocation and hierarchical classification efficiency; a separate LSE must be used for each criterion.

composite score[a] --A score that combines several scores by a specified formula.

differential assignment theory (DAT)--DAT is defined by four organizing concepts: (1) to maximize benefits, a set of quantitative principles must be employed that embrace selection of predictors in a battery, the structure of job families and the strategies and algorithms used in the selection/assignment process; (2) utility models, measuring benefits in terms of mean predicted performance, provide the best approach for specifying personnel selection policies and procedures for operational systems; (3) benefits for both selection and classification procedures are maximized by using the same weights for a given set of composites under optimal conditions, while under non-optimal conditions, selection and classification must be separately considered; and (4) any multidimensional selection/classification strategy and algorithm can be practically implemented in operations by utilizing available computer capabilities.

differential validity--The level of prediction using LSEs of differences among criterion scores when referring to H_d; this measure is related to the variation of a validity vector with jobs and to an assignment variable being more valid for its own job family than any other job family.

efficiency--A solution that minimizes costs as measured by physical resources and time utilized.

hierarchical classification efficiency--All classification efficiency not explainable as allocation efficiency; it capitalizes on disparate variances of the mean predicted benefit scores for the corresponding jobs.

hierarchical layering--A phenomenon in which LSEs are more valid or of more value for some jobs than for others.

mean predicted performance (MPP)--The measurement of benefits can be approximated by computing MPP across jobs; if MPP is weighted by the value of each job, it becomes a more useful measure of benefits. It provides a means of comparing the effectiveness of alternative sets of tests or test batteries in the context of a specified set of jobs and performance scores.

multidimensional screening (MDS)--A selection/classification process using an algorithm that ensures that no non-selected person has a higher predicted performance on any job than the person assigned to that job; the algorithm also ensures that no other assignment can further raise the mean predicted performance.

norm-referenced test[a] --An instrument for which interpretation is based on the comparison of a test taker's performance to the performance of other people in a specified group.

operational efficiency--The improvement in MPP obtained from the usually imperfect operational selection assignment process as contrasted to potential efficiency; the improvement obtainable if the maximally efficient prediction composites of a given battery were to be used in optimal selection/assignment decisions.

potential allocation efficiency--The maximum allocation effectiveness achievable from the differential validity of a given test battery and set of jobs expressed as a mean predicted performance standard score.

potential classification efficiency--The maximum classification effectiveness achievable from a given test battery and set of jobs expressed as a mean predicted performance standard score; it incorporates both potential allocation and hierarchical layering effects.

potential selection efficiency--Rank-ordering applicants on some benefit continuum and rejecting all those below some point on that continuum.

predictor[a] --A measurable characteristic that predicts criterion performance such as scores on a test, evidence of previous performance, and judgments of interviewers, panels or raters.

psychometric[a] --Pertaining to the measurement of psychological characteristics such as abilities, aptitudes, achievement, personality, traits, skill and knowledge.

regression equation[a] --An algebraic equation used to predict criterion performance from predictor scores.

reliability[a] --The degree to which test scores are consistent, dependable or repeatable; that is, the degree to which they are free of errors of measurement.

restriction of range[a] --A situation in which, because of sampling restrictions, the variability of data in the sample is less than the variability in the population of interest.

sample[b] --The individuals who are actually tested from among those in the population to which the procedure is to be applied.

score[a] --Any specific number resulting from the assessment of an individual; a generic term applied for convenience to such diverse measures as test scores, estimates of latent variables, production counts, absence records, course grades, ratings, and so forth.

selection--A procedure for rejecting some applicants for organizational membership as contrasted to assigning all applicants to jobs (classification); or rejecting an applicant for a single job as contrasted to selection and assignment to one of a number of jobs (multidimensional selection).

simulation model[c] --A special type of abstract model that is analogous to a segment of the real world and contains a time dimension. It is used to explain and predict behavior as if it occurred on the real world.

standard score[a] --A score that describes the location of a person's score within a set of scores in terms of its distance from the mean in standard deviation units.

test[b] --A measure based on a sample of behavior.

unidimensionality[a] --A characteristic of a test that measures only one latent variable.

utility[c] --Technically, want-satisfying power; it is often defined as the preference of the decisionmaker for a given outcome.

utility analysis--The determination of institutional gain or loss (outcomes) anticipated from various courses of action usually measured in terms of dollars.

validity[a] --The degree to which a certain inference from a test is appropriate or meaningful.

validity coefficient[a] --A coefficient of correlation that shows the strength of the relation between the predictor and criterion.

validity generalization--Often used synonomously with a particular approach to meta-analysis of validity data across studies. Our use of the term is also intended to reflect the stability of validity findings for all ability measures across jobs and situations.

variable[a] --A quantity that may take on any one of a specified set of values.

NOTES.

a Adapted from American Psychological Association. American Educational Research Association, and National Council on Measurement in Education (1985), *Standards for Education and Psychological Testing.*

b Adapted from Society for Industrial and Organizational Psychology (1987), *Principles for the Validation and Use of Personnel Selection Procedures.*

c Adapted from Heyne (1988), *Microeconomics.*

REFERENCES

Brogden, H. (1955). Least squares estimates and optimal classification. *Psychometrika, 20(3)*, 249-252.

Brogden, H. (1959). Efficiency of classification as a function of number of jobs, per cent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement, 19(2)*, 181-190.

Campbell, J.P. (Ed.). (1987). *Improving the selection, classification and utilization of Army enlisted personnel: Annual report, 1985 fiscal year* (Technical Report 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cronbach, L.J. and Gleser, G.C. (1965). *Psychological tests and personnel decisions.* (2nd edition). Urbana: University of Illinois Press.

Dwyer, P.S. (1954). Solution of the personnel classification problem with the method of optimal regions. *Psychometrika, 18(1)*, 11-26.

Fishman, G.S. and Moore, L.R. (1986). An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal of Science and Statistical Computing, 7(1)*, 24-45.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Harris, R.N. (1967). *A model sampling experiment to evaluate two methods of test selection* (Research Memorandum 67-2). Washington, D.C.: U.S. Army Behavioral Science Research Laboratory.

Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychometrika, 68(9)*, Whole No. 380.

Horst, P. (1956). Multiple classification by the method of least squares. *Journal of Clinical Psychology, 12*, 3-16.

Horst, P. and MacEwan, C. (1960). Predictor elimination techniques for determining multiple prediction batteries, *Psychological Reports, Monograph Supplement 1-V7, 7*, 19-50.

Hough, L. and Ashworth, S. (1986). *Concurrent validity data analyses non-cognitive predictors (trial battery).* Personnel Decisions Research Institute.

Hunter, J.E., Crosson, I.J., and Friedman, D.H. (1985). *The validity of the Armed Services Vocational Aptitude Battery for civilian and military job performance.* Rockville, MD: Research Applications Inc.

Johnson, C.D. and Sorenson, R.C. (1974). Model sampling experimentation for manpower planning. In Clough, D.J., Lewis, C.C., and Oliver, A.L. (Eds.). *Manpower Planning Models*, London: English University Press (pp. 33-52).

Johnson, C.D., Klieger, W.A., and Frankfeld*, E. (1958, April). *Construction of a classification self-description blank designed to emphasize differences among occupational areas*. Research Memorandum 58-6. Personnel Research Branch, The Adjutant General's Office, Dept. of the Army: Washington, D.C.

Johnson, C.D. and Kotula, L.J. (1958, August). *Validation of experimental self-description materials for general and differential classification*. Technical Research Note 95. Personnel Research Branch, The Adjutant General's Office, Dept. of the Army: Washington, D.C.

Johnson, C.D. and Zeidner, J. (April, 1990). *Classification Utility: Measuring and Improving Benefits in Matching Personnel to Jobs* (IDA Paper P-2240). Alexandria: VA. Institute for Defense Analyses.

Kennington, J,L. and Helgason, R.V. (1980). *Algorithms for network programming*. John Wiley and Sons.

Lawley, D. (1943). A note on Karl Pearson's selection formulae. *Royal Society of Edinburgh Proceedings, Section A, 62*, 28-30.

Lord, F.M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Menlo Park, California: Addison-Wesley.

Maier, M.H. and Grafton, F.C. (1981). *Aptitude composites for ASVAB 8, 9 and 10* (ARI Research Report No. 1308). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McLaughlin, D.H., Rossmeissl, P.G., Wise, L.L., Brandt, D.A., and Wang, M. (1984). *Project A: Validation of current and alternative ASVAB area composites, based on training and SQT information on FY 1981 and FY 1982 enlisted accessions* (ARI Technical Report 651). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mitchell, K.J. and Hanser, L.M. (1984). *The 1980 Youth Population Norms: Enlistment and Occupational Classification Standards in the Army*. Technical Report. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Niehl, E. (1967). *A General Computer Simulation for Conducting Allocation Experiments* (Research Memorandum 67-1), Washington, D.C.: U.S. Army Personnel Research Office. (ADA079330).

Niehl, E. and Sorenson, R.C. (1968). *Simpo-I entity model for determining the qualitative impact on personnel policies* (Technical Research Note, 193). Washington D.C.: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD831268).

Nord, R. and Schmitz, E. (September, 1989). Estimating performance and utility effects of alternative selection and classification policies. In Zeidner and Johnson (September 1989). *The economic benefits of predicting job performance* (IDA Paper P-2241). Alexandria, VA: Institute for Defense Analyses.

Nord, R. and White, L.A. (1988). The mesurement and application of performance utility: Some key issues. In B.F. Green, Jr., H. Wing, and A.K. Wigdor (Eds.), *Linking military enlistment standards to job performance: Report of a workshop* (pp. 215- 243). Washington, D.C.: National Academy Press.

Olson, P.T., Sorenson, R.C., Haynman, K.W., Witt, J.M., and Abbe, E.N. (1969). *Summary of SIMPO-I model development* (Technical Research Report 1157). Washington, D.C.: U.S. Army Behavioral Science Research Laboratory.

Park, S.K. and Miller, K.W. (1988). Random number generators: Good ones are hard to find. *Communications of the ACM, 31(10)*, 1192-1201.

Sadacca, R., Campbell, J.P., White, L.A., and DiFazio, A. (1988). *Weighting criterion components to develop composite measures of job performance.* HumRRO IR-PRD-88-14.

Schmidt, F.L., Hunter, J.E., and Dunn, W.L. (1987). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery (ASVAB).* San Diego, CA: Naval Personnel Research and Development Center.

Schmidt, F.L., Hunter, J.E., and Larson, M. (1988). *General cognitive ability vs. general and specific aptitudes in the prediction of training performance: Some preliminary findings.* Navy Personnel Research and Development Center: San Diego, CA.

Sorenson, R.C. (1965). *Optimal allocation of enlisted men--Full regression equations versus aptitude area scores* (Technical Research Note 163). Washington, D.C.: U.S. Army Personnel Research Office. (AD625224).

Thorndike, R.L. (1950). The problem of classification of personnel. *Psychometrika, 15*, 215-235.

Wise, L.L., Campbell, J.P., and Peterson, N.G. (1987, April). *Identifying optimal predictor composites and testing for generalizability across jobs and performance constructs.* Paper presented at the Mid-year Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Wise, L.L., Campbell, J.P., McHenry, J.J., and Hanser, L.R. (1986, August). *A latent structure model of job performance factors.* Paper presented at the American Psychological Association Convention, Washington, D.C.

Zeidner, J. (April, 1987). *The validity of selection and classification procedures for predicting job performance* (IDA Paper P-1977). Alexandria, VA: Institute for Defense Analyses.

Zeidner, J. and Johnson, C. (July, 1989a). *The utility of selection for military and civilian jobs.* (IDA Paper P-2239). Alexandria, VA: Institute for Defense Analyses.

Zeidner, J. and Johnson, C. (September, 1989b). *The economic benefits of predicting job performance.* (IDA Paper P-2241). Alexandria, VA: Institute for Defense Analyses.

APPENDIX A

JOB SAMPLE

TABLE A-3: MOS DESCRIPTIONS

11B	Leads, supervises and serves as a member of an infantry activity involving machine guns and other weapons in offensive and defensive combat operations. Duty is to destroy enemy personnel weapons and equipment.
12B	Commands, serves or assists as a member of a team, squad, section or platoon engaged in providing combat engineering support to combat forces. Performs combat construction, demolitions and related duties.
13B	Supervises or serves as a member of field artillery cannon unit. Participates in emplacement, laying, firing and displacement of field artillery cannons.
16S	Supervises or serves as a member of MANPADS missile unit by preparing and firing MANPADS missile.
19E	Leads, supervises or serves as a member of M49-M60 armor unit in offensive and defensive combat operations. Loads and fires tank main gun.
27E	Supervises or performs direct and general support maintenance on the TOW and DRAGON missile systems, trainers, nightsights, battery chargers and system peculiar test and check-out equipment.
31C	Operates single channel radio, radio teletype and satellite equipment.
51B	Performs general and heavy carpentry and masonry duties in fabrication, erection, maintenance and repair of wooden and masonry structures and articles.
54E	Performs NBC reconnaissance and operates and maintains identification/detection and decontamination equipment.
55B	Supervises, performs or assists in storage, receipt issue, stock control, accounting and maintenance operations involving ammunition, ammunition components and explosives.
63_	Performs and supervises organizational maintenance and recovery operations on light wheel vehicles (prime movers designated as five to or less and their associated trailers). Troubleshoots and performs unit maintenance on internal combustion engines and accessories, powertrain and chassis components.
64C	Supervises or operates wheel vehicles to transport personnel and cargo.
67N	Supervises, inspects or performs maintenance on .ility helicopters, excluding repair of systems components. Assists in organizational, direct and general support (aviation unit, intermediate and depot) maintenance of utility helicopters.
71L	Supervises or performs administrative, clerical and typing duties.
76W	Supervises or receives, stores, account and cares for, dispenses, issues and ships bulk or packaged petroleum, oils and lubricants (POL) products.
76Y	Supervises or performs duties involving request, receipt, storage issue, accounting for and preservation of individual, organizational, installation, and expendable supplies and equipment in a unit.
91A	Supervises dispensary or field medical facilities, administers emergency medical treatment to battlefield casualties, assists with in-patient and out-patient care and treatment, and assists with technical and administrative management of medical treatment facilities.
94B	Supervises or prepares and cooks food in field, garrison or central food preparation activities.
95B	Sup..vises or provides law enforcement activities, preserves military control, controls traffic, quells disturbances, protects property and personnel, handles prisoners of war, refugees or evacuees, and investigates incidents.

Source: adapted from Hough and Ashworth (1986)

APPENDIX B

PREDICTOR MEASURES

Table B-1. ASVAB/Project A Composites and Reliabilities

Code	Predictor/Composite	Reliability[a]
	ASVAB subtests	
GS	General Science	0.86
AR	Arithmetic Reasoning	0.91
NO	Numerical Operations	0.78
CS	Coding Speed	0.85
AS	Auto Shop Information	0.87
MK	Mathematical Knowledge	0.87
MC	Mechanical Comprehension	0.85
EI	Electronics Information	0.82
PC	Paragraph Comprehension	0.81
WK	Word Knowledge	0.92
	Paper-and-pencil spatial composite[b]	
SPAT	Spatial composite	0.71
	Perceptual-psychomotor composites[c]	
CPAC	Complex perceptual accuracy composite	0.62
CPSP	Complex perceptual speed composite	0.95
NMSA	Number speed and accuracy	0.84
PSYM	Psychomotor composite	0.82
SRAC	Simple reaction accuracy composite	0.52
SRSP	Simple reaction speed composite	0.88
	Job orientation composites (JOB)[d]	
AUTO	Autonomy composite	0.50
SUPP	Organizational & Co-worker Support	0.67
ROUT	Routine composite	0.46
	Temperament and biodata composites (ABLE)[e]	
ADJU	Adjustment composite	0.74
DEPN	Dependability composite	0.76
COND	Physical condition composite	0.85
SURG	Achievement orientation composite	0.78
	Interest composites (AVOICE)[e]	
AUDI	Audiovisual interest composite	0.74
COMB	Combat interest composite	0.78
FSER	Food service interest composite	0.67
PSER	Protective service interest composite	0.76
TECH	Technical interest composite	0.75
MACH	Machinery interest composite	0.79

Note:
[a] ASVAB reliabilities reported in McLaughlin, et al. (1984), p. 9, Project A reliabilities reported in 1985 Report.
[b] Test-retest reliability (N = 468 to 487).
[c] Odd-even reliability.
[d] Internal consistency reliability (alpha).
[e] Test-retest reliabilities (N=389 to 412).

TABLE B-2: ARMY COMPOSITES COMPUTED FROM ASVAB TEST SCORES

Code	Composite	ASVAB Test Formula
AFQT	Armed Forces Qualification Test	AR + NO/2 + VE
CLER	Clerical	VE + NO + CS
CO	Combat	CS + AR + MC + AS
ELEC	Electronics	AR + MK + EI + GS
FA	Field Artillery	CS + AR + MC + MK
GM	General Maintenance	MK + EI + GS + AS
OF	Operators and Food	NO + VE + MC + AS
MM	Mechanical Maintenance	NO + EI + MC + AS
SC	Surveillance and Communications	NO + CS + VE + AS
ST	Skilled Technician	VE + MK + MC + GS

Source: Maier & Grafton (1981)
Note. The AA composites were normed against the 1980 Youth Reference Population to have a mean of 100 and standard deviation of 20

TABLE B-3: PROJECT A PREDICTOR CONSTRUCTS & MEASURES

Construct	Composites, Tests & Scales
Spatial Visualization & Orientation	Cognitive paper-and-pencil Assembling Objects; Object Rotation; Maze Test; Path Test; Orientation Test
Perceptual/Psychomotor	Computer-Administered Complex perceptual speed; Complex perceptual accuracy; Number memory; Target Identification; Target Tracking; Target Shoot; Cannon Shoot; Simple Reaction Time; Choice Reaction Time
Temperament/Personality	Assessment of Background & Life Experiences (ABLE) Adjustment; Dependability; Achievement Orientation; Physical condition
Vocational Interest	Army Vocational Interest Career Examination (AVOICE) Audiovisual interest; Combat interest; Protective service; Machinery interest; Food service interest; Technical interest
Job Reward	Job Orientation Blank (JOB) Organizational & co-worker support; Autonomy; Routine work

Source: adapted from Campbell (1987)

APPENDIX C

CRITERION MEASURES

APPENDIX C CRITERION MEASURES

TABLE C-1: PROJECT A CRITERION COMPONENTS AND RELIABILITIES

Code	Criterion Component	Reliability[a]
CTP	MOS specific core technical skills	0.85
GSP	General skills proficiency	0.85
ELS	Effort, Leadership and Self-development	0.80
MPD	Personal Discipline	0.80
PFB	Physical Fitness & Military Bearing	0.80

[a] Source: Zeidner (1987)

TABLE C-2: DESCRIPTIONS OF CRITERION MEASURES

Code	Description
CTP	Individual proficiency in tasks which are central to MOS, i.e. core tasks which are a job's primary definers across MOS. **Measured by:** A total, standardized score from the within method sum of core content job knowledge and job sample tests.
GSP	Individual proficiency in general or common tasks across MOS, e.g., first aid, use of basic weapons, etc. **Measured by:** A total, standardized factor score from the within-method sum of general task scales.
ELS	The degree to which the individual exerts effort over the full range of job tasks, perseveres under adverse or dangerous conditions, and demonstrates leadership or support towards peers. Assesses general dependability, judgment and perseverance. **Measured by:** Five scales from the Army-wide Behaviorally Anchored Rating Scale (BARS) (general technical performance, leadership, effort, self- development and general maintenance); expected combat performance scales; MOS-specific BARS scales; and the total of commendations/awards.
MPD	Adherence to Army regulations and traditions, personal self- control, responsibility in day-to-day Behavior, and personal discipline. **Measured by:** Three Army-wide BARS scales (adherence to traditions and regulations, self control, integrity); combat rating scale; and two indices from administrative records (no. of disciplinary actions & promotion rate).
PFB	Appropriate military appearance and good physical condition. **Measured by:** Physical fitness qualification score from personnel records and "military bearing and appearance" rating scale.

Source: Descriptions adapted from Campbell (1987)

TABLE C-3: MEAN CRITERION CONSTRUCT WEIGHTS BY OFFICERS FOR 19 MOS USING THE CONJOINT METHOD

MOS	General Skills	MOS Skills	Leadership & Effort	Personal Discipline	Military Bearing
11B	18.5	22.9	29.1	17.2	12.3
12B	19.7	18.4	30.2	20.3	11.5
13B	19.2	22.7	27.7	18.3	12.1
16S	16.3	25.9	26.3	20.2	11.4
19E	21.1	29.4	20.5	19.5	11.0
27E	18.0	24.2	22.4	23.0	12.4
31C	20.3	29.0	22.0	17.3	11.4
51B	17.2	25.6	25.6	19.7	11.9
54E	21.5	25.4	20.7	19.8	12.6
55B	19.5	22.4	27.8	19.5	10.8
63B	18.1	27.5	23.5	21.1	9.9
64C	22.8	26.1	21.8	15.4	14.0
67N	15.9	25.9	25.3	22.2	10.6
71L	19.9	24.1	22.7	21.0	12.3
76W	17.2	23.6	25.0	22.9	11.4
76Y	21.7	25.7	19.8	17.5	15.3
91A	16.6	26.9	23.1	22.5	11.0
94B	17.4	24.5	26.2	20.3	11.0
95B	27.8	20.0	20.5	19.1	12.6

Source: Saddaca, Campbell, White and DiFazio (1988)

APPENDIX D

POPULATION DATA

TABLE D-1: 1980 YOUTH POPULATION ASVAB INTERCORRELATIONS (see Appendix B for code names)

	GS	AR	NO	CS	AS	MK	MC	EI	VE
GS	1.00	.72	.52	.45	.64	.69	.70	.76	.80
AR	.72	1.00	.63	.51	.53	.83	.69	.66	.73
NO	.52	.63	1.00	.70	.30	.62	.40	.41	.62
CS	.45	.51	.70	1.00	.22	.52	.34	.34	.57
AS	.64	.53	.30	.22	1.00	.41	.74	.75	.52
MK	.69	.83	.62	.52	.41	1.00	.60	.59	.70
MC	.70	.69	.40	.34	.74	.60	1.00	.74	.60
EI	.76	.66	.41	.34	.75	.59	.74	1.00	.67
VE	.80	.73	.62	.57	.52	.70	.60	.67	1.00

Source: Army Research Institute, Project A, 13 July 1988

TABLE D-2: POPULATION PREDICTOR INTERCORRELATIONS (see Appendix B for code names)

PREDICTORS 1-8

	GS	AR	NO	CS	AS	MK	MC	EI
GS	1.0000	0.7200	0.5200	0.4500	0.6400	0.6900	0.7000	0.7600
AR	0.7200	1.0000	0.6300	0.5100	0.5300	0.8300	0.6900	0.6600
NO	0.5200	0.6300	1.0000	0.7000	0.3000	0.6200	0.4000	0.4100
CS	0.4500	0.5100	0.7000	1.0000	0.2200	0.5200	0.3400	0.3400
AS	0.6400	0.5300	0.3000	0.2200	1.0000	0.4100	0.7400	0.7500
MK	0.6900	0.8300	0.6200	0.5200	0.4100	1.0000	0.6000	0.5900
MC	0.7000	0.6900	0.4000	0.3400	0.7400	0.6000	1.0000	0.7400
EI	0.7600	0.6600	0.4100	0.3400	0.7500	0.5900	0.7400	1.0000
VE	0.8000	0.7300	0.6200	0.5700	0.5200	0.7000	0.6000	0.6700
SPAT	0.6707	0.7301	0.5162	0.4877	0.5677	0.6802	0.7413	0.6159
CPAC	0.3166	0.3560	0.3047	0.3155	0.2084	0.3485	0.2775	0.2749
CPSP	0.3170	0.2876	0.3119	0.2953	0.2427	0.2811	0.3005	0.2617
NMSA	0.5895	0.7156	0.6966	0.5545	0.3938	0.6774	0.4914	0.4996
PSYM	0.4459	0.4383	0.3249	0.2920	0.4586	0.3841	0.5479	0.4544
SRAC	0.2136	0.2179	0.1653	0.1703	0.1901	0.1861	0.2100	0.2023
SRSP	0.2169	0.2283	0.2646	0.2534	0.1385	0.2146	0.1892	0.1766
AUTO	0.2486	0.2261	0.1849	0.1562	0.2227	0.1953	0.2203	0.2393
SUPP	0.1383	0.1196	0.1739	0.1745	0.0436	0.1294	0.0584	0.0938
ROUT	-0.3150	-0.3021	-0.2525	-0.2355	-0.2507	-0.2620	-0.2898	-0.2737
ADJU	0.2256	0.2399	0.1925	0.1338	0.2048	0.2147	0.2261	0.2259
DEPN	0.0522	0.1017	0.1350	0.1520	-0.0384	0.1450	0.0162	0.0330
COND	-0.0462	-0.0322	-0.0048	-0.0387	-0.0147	-0.0269	-0.0123	-0.0348
SURG	0.2076	0.2533	0.2371	0.2020	0.1593	0.2393	0.1903	0.2003
AUDI	0.0147	0.0022	0.0058	0.0221	-0.0909	0.0482	-0.0184	-0.0171
COMB	0.1539	0.0660	-0.0309	-0.0663	0.3433	0.0120	0.2594	0.2220
FSER	-0.2097	-0.1852	-0.1295	-0.1199	-0.2366	-0.1408	-0.2317	-0.2179
PSER	-0.0990	-0.1426	-0.1365	-0.1275	0.0101	-0.1601	-0.0577	-0.0580
TECH	-0.0039	0.0629	0.1116	0.0783	-0.1353	0.1275	-0.0575	-0.0342
MACH	-0.1545	-0.1908	-0.2822	-0.2951	0.1864	-0.2210	0.0465	0.0075

TABLE D-2 (CONT.): POPULATION PREDICTOR INTERCORRELATIONS

PREDICTORS 9-16

	VE	SPAT	CPAC	CPSP	NHSA	PSYM	SRAC	SRSP
GS	0.8000	0.6707	0.3166	0.3170	0.5895	0.4459	0.2136	0.2169
AR	0.7300	0.7301	0.3560	0.2876	0.7156	0.4383	0.2179	0.2283
NO	0.6200	0.5162	0.3047	0.3119	0.5966	0.3249	0.1653	0.2646
CS	0.5700	0.4877	0.3155	0.2953	0.5545	0.2920	0.1703	0.2534
AS	0.5200	0.5677	0.2084	0.2427	0.3938	0.4586	0.1901	0.1385
MK	0.7000	0.6802	0.3485	0.2811	0.6774	0.3841	0.1861	0.2146
MC	0.6000	0.7413	0.2775	0.3005	0.4914	0.5479	0.2100	0.1892
EI	0.6700	0.6159	0.2749	0.2617	0.4996	0.4544	0.2023	0.1766
VE	1.0000	0.6234	0.3678	0.2783	0.6498	0.3773	0.2312	0.2288
SPAT	0.6234	1.0000	0.3886	0.4057	0.6143	0.6040	0.2311	0.2569
CPAC	0.3678	0.3886	1.0000	-0.2025	0.3000	0.2477	0.2284	0.0695
CPSP	0.2783	0.4057	-0.2025	1.0000	0.4129	0.3768	0.0642	0.3716
NHSA	0.6498	0.6143	0.3000	0.4129	1.0000	0.4413	0.1983	0.3023
PSYM	0.3773	0.6040	0.2477	0.3768	0.4413	1.0000	0.1434	0.2696
SRAC	0.2312	0.2311	0.2284	0.0642	0.1983	0.1434	1.0000	0.1200
SRSP	0.2288	0.2569	0.0695	0.3716	0.3023	0.2696	0.1200	1.0000
AUTO	0.2602	0.2039	0.0566	0.1009	0.1890	0.1371	0.0368	0.0681
SUPP	0.2090	0.0978	0.0886	0.0516	0.1497	0.0528	0.0561	0.0635
ROUT	-0.3420	-0.2974	-0.1429	-0.1408	-0.2577	-0.2091	-0.0919	-0.1225
ADJU	0.2315	0.2258	0.1227	0.1186	0.2093	0.1934	0.0694	0.1149
DEPN	0.0889	0.0561	0.1070	-0.0034	0.0990	-0.0230	0.0245	0.0250
COND	-0.0556	-0.0352	-0.0547	0.0688	0.0062	0.0990	-0.0456	0.0477
SURG	0.2392	0.2023	0.1246	0.0997	0.2301	0.1352	0.0498	0.0991
AUDI	0.0507	0.0032	0.0199	0.0052	-0.0293	-0.0224	-0.0241	-0.0152
COMB	0.0435	0.1737	0.0135	0.0728	0.0276	0.2522	0.0092	0.0196
FSER	-0.1939	-0.2148	-0.0924	-0.1278	-0.1650	-0.2282	-0.0718	-0.1014
PSER	-0.1356	-0.0907	-0.0818	0.0008	-0.1145	0.0214	-0.0293	-0.0071
TECH	0.0483	-0.0134	0.0601	-0.0156	0.0688	-0.0316	-0.0292	0.0116
MACH	-0.2955	-0.0620	-0.1171	-0.0538	-0.2163	0.0616	-0.0653	-0.0809

TABLE D-2 (CONT.)

PREDICTORS 17-24

	AUTO	SUPP	ROUT	ADJU	DEPN	COND	SURG	AUDI
GS	0.2486	0.1383	-0.3150	0.2256	0.0522	-0.0462	0.2076	0.0147
AR	0.2261	0.1196	-0.3021	0.2399	0.1017	-0.0322	0.2533	0.0022
NO	0.1849	0.1739	-0.2525	0.1925	0.1350	-0.0048	0.2371	0.0058
CS	0.1562	0.1745	-0.2355	0.1338	0.1520	-0.0387	0.2020	0.0221
AS	0.2227	0.0436	-0.2507	0.2048	-0.0384	-0.0147	0.1593	-0.0909
MK	0.1953	0.1294	-0.2620	0.2147	0.1450	-0.0269	0.2393	0.0482
MC	0.2203	0.0584	-0.2898	0.2261	0.0162	-0.0123	0.1903	-0.0184
EI	0.2393	0.0938	-0.2737	0.2259	0.0330	-0.0348	0.2003	-0.0171
VE	0.2602	0.2090	-0.3420	0.2315	0.0889	-0.0556	0.2392	0.0507
SPAT	0.2039	0.0978	-0.2974	0.2258	0.0561	-0.0352	0.2023	0.0032
CPAC	0.0566	0.0886	-0.1429	0.1227	0.1070	-0.0547	0.1246	0.0199
CPSP	0.1009	0.0516	-0.1408	0.1186	-0.0034	0.0688	0.0997	0.0052
NHSA	0.1890	0.1497	-0.2577	0.2093	0.0990	0.0062	0.2301	-0.0293
PSYM	0.1371	0.0528	-0.2091	0.1934	-0.0230	0.0990	0.1352	-0.0224
SRAC	0.0368	0.0561	-0.0919	0.0694	0.0245	-0.0456	0.0498	-0.0241
SRSP	0.0681	0.0635	-0.1225	0.1149	0.0250	0.0477	0.0991	-0.0152
AUTO	1.0000	0.2877	-0.1530	0.1069	0.0051	0.0531	0.2010	0.1033
SUPP	0.2877	1.0000	-0.2384	0.1163	0.2542	0.0584	0.3358	0.1682
ROUT	-0.1530	-0.2384	1.0000	-0.1912	-0.0363	-0.0653	-0.2435	-0.0059
ADJU	0.1069	0.1163	-0.1912	1.0000	0.3414	0.2268	0.6038	0.0622
DEPN	0.0051	0.2542	-0.0363	0.3414	1.0000	0.1279	0.5971	0.1924
COND	0.0531	0.0584	-0.0653	0.2268	0.1279	1.0000	0.3410	0.0622
SURG	0.2010	0.3358	-0.2435	0.6038	0.5971	0.3410	1.0000	0.1838
AUDI	0.1033	0.1682	-0.0059	0.0622	0.1924	0.0622	0.1838	1.0000
COMB	0.1373	0.0294	-0.0808	0.1666	-0.0298	0.1537	0.1868	0.1781
FSER	-0.0944	-0.0527	0.2245	-0.0707	0.0489	-0.0341	-0.0412	0.3074
PSER	0.0002	0.0635	0.0482	0.0392	0.0340	0.1304	0.0790	0.1378
TECH	0.0684	0.2415	0.0084	0.1489	0.3069	0.0869	0.2955	0.6719
MACH	0.0138	-0.0697	0.1119	0.0014	-0.1022	0.1296	0.0076	0.2014

TABLE D-2 (CONT.): POPULATION PREDICTOR INTERCORRELATIONS

PREDICTORS 25-29

	COMB	FSER	PSER	TECH	MACH
GS	0.1539	-0.2097	-0.0990	-0.0039	-0.1545
AR	0.0660	-0.1852	-0.1426	0.0629	-0.1908
NO	-0.0309	-0.1295	-0.1365	0.1116	-0.2822
CS	-0.0663	-0.1199	-0.1275	0.0783	-0.2951
AS	0.3433	-0.2366	0.0101	-0.1353	0.1864
MK	0.0120	-0.1408	-0.1601	0.1275	-0.2210
MC	0.2594	-0.2317	-0.0577	-0.0575	0.0465
EI	0.2220	-0.2179	-0.0580	-0.0342	0.0075
VE	0.0435	-0.1939	-0.1356	0.0483	-0.2955
SPAT	0.1737	-0.2148	-0.0907	-0.0134	-0.0620
CPAC	0.0135	-0.0924	-0.0818	0.0601	-0.1171
CPSP	0.0728	-0.1278	0.0008	-0.0156	-0.0538
NMSA	0.0276	-0.1650	-0.1145	0.0688	-0.2163
PSYM	0.2522	-0.2282	0.0214	-0.0316	0.0616
SRAC	0.0092	-0.0718	-0.0293	-0.0292	-0.0653
SRSP	0.0196	-0.1014	-0.0071	0.0116	-0.0809
AUTO	0.1373	-0.0944	0.0002	0.0684	0.0138
SUPP	0.0294	-0.0527	0.0635	0.2415	-0.0697
ROUT	-0.0808	0.2245	0.0482	0.0084	0.1119
ADJU	0.1666	-0.0707	0.0392	0.1489	0.0014
DEPN	-0.0298	0.0489	0.0340	0.3069	-0.1022
COND	0.1537	-0.0341	0.1304	0.0869	0.1296
SURG	0.1868	-0.0412	0.0790	0.2955	0.0076
AUDI	0.1781	0.3074	0.1378	0.6719	0.2014
COMB	1.0000	0.0864	0.3913	0.1905	0.5881
FSER	0.0864	1.0000	0.1708	0.3518	0.2269
PSER	0.3913	0.1708	1.0000	0.2216	0.3364
TECH	0.1905	0.3518	0.2216	1.0000	0.2118
MACH	0.5881	0.2269	0.3364	0.2118	1.0000

TABLE D-3: PREDICTOR VALIDITY COEFFICIENTS FOR 19 MOS USING SINGLE (CTP) CRITERION

PREDICTORS 1-8

MOS	GS	AR	NO	CS	AS	MK	MC	EI
1	0.6474	0.6285	0.5424	0.4772	0.4808	0.6309	0.5561	0.5711
2	0.6895	0.6255	0.4760	0.3488	0.5842	0.6285	0.6265	0.6536
3	0.4179	0.3814	0.3386	0.2657	0.4336	0.3251	0.3897	0.3833
4	0.4960	0.5481	0.4003	0.3978	0.3307	0.5517	0.4020	0.3993
5	0.6123	0.5318	0.4251	0.3451	0.4856	0.5476	0.5301	0.5492
6	0.6587	0.6195	0.6300	0.5674	0.5174	0.5688	0.5604	0.6152
7	0.4723	0.5421	0.3194	0.1876	0.3727	0.5271	0.4452	0.4761
8	0.7057	0.6665	0.726[illegible]	0.6146	0.5683	0.6565	0.6588	0.5572
9	0.5859	0.6141	0.4297	0.3863	0.4947	0.5936	0.5133	0.5603
10	0.4819	0.4067	0.3632	0.3585	0.3731	0.3643	0.4531	0.4325
11	0.4660	0.4143	0.2207	0.2500	0.5822	0.3545	0.5521	0.5134
12	0.2972	0.3480	0.0885	0.0819	0.3786	0.3209	0.4033	0.3495
13	0.3958	0.3876	0.2425	0.2096	0.3495	0.3895	0.4042	0.4161
14	0.4517	0.5286	0.3734	0.3276	0.2269	0.5674	0.3453	0.3253
15	0.7178	0.6852	0.4269	0.4707	0.7020	0.6392	0.7028	0.6818
16	0.5710	0.6108	0.4197	0.4012	0.4416	0.6335	0.4854	0.5559
17	0.4682	0.4395	0.3757	0.4378	0.3897	0.4701	0.4124	0.4189
18	0.5546	0.6800	0.5202	0.5034	0.3976	0.6157	0.4891	0.4755
19	0.3057	0.3562	0.3345	0.2795	0.2316	0.3607	0.2710	0.3117

TABLE D-3 (CONT.): PREDICTOR VALIDITY COEFFICIENTS FOR 19 MOS USING CTP CRITERION

PREDICTORS 9-16

MOS	VE	SPAT	CPAC	CPSP	NMSA	PSYM	SRAC	SRSP
1	0.6361	0.6646	0.3630	0.3415	0.5357	0.4177	0.2229	0.2483
2	0.6584	0.6217	0.3127	0.2198	0.5008	0.3574	0.1782	0.1816
3	0.4072	0.4817	0.2669	0.2163	0.3855	0.3377	0.1300	0.2321
4	0.4864	0.5444	0.3132	0.1611	0.4858	0.4078	0.0871	0.1181
5	0.5398	0.5841	0.3791	0.2182	0.5249	0.3569	0.2296	0.1966
6	0.6547	0.5230	0.3041	0.2643	0.5701	0.4099	0.1813	0.1580
7	0.4437	0.4493	0.2642	0.1076	0.4203	0.2766	0.0926	0.1154
8	0.7119	0.7750	0.5240	0.1402	0.6348	0.4727	0.1465	0.2465
9	0.5425	0.5921	0.3127	0.2333	0.5284	0.3685	0.2311	0.1295
10	0.5438	0.5080	0.3242	0.1341	0.3653	0.4061	0.2264	0.1407
11	0.3755	0.5238	0.1494	0.2356	0.3101	0.3578	0.1643	0.1850
12	0.2284	0.4006	0.2394	0.1842	0.2639	0.2504	0.1542	0.0858
13	0.3595	0.4551	0.2167	0.0433	0.2940	0.2838	0.0930	0.0473
14	0.4752	0.5188	0.3649	0.1403	0.4078	0.2740	0.1775	0.1540
15	0.6940	0.7169	0.3085	0.3548	0.6012	0.5001	0.3682	0.1713
16	0.5772	0.5423	0.3185	0.1671	0.5631	0.2525	0.2379	0.2133
17	0.4361	0.5034	0.2270	0.1912	0.4151	0.3101	0.1525	0.1563
18	0.6117	0.6485	0.4664	0.2430	0.6001	0.3161	0.3066	0.2250
19	0.3283	0.3737	0.2396	0.1293	0.3444	0.2507	0.1044	[illegible]

TABLE D-3 (CONT.)

PREDICTORS 17-24

MOS	AUTO	SUPP	ROUT	ADJU	DEPN	COND	SURG	AUSI
1	0.2029	0.1417	-0.2976	0.2241	0.1675	0.0355	0.2971	0.0138
2	0.2236	0.1086	-0.3019	0.1965	0.0468	-0.0279	0.2150	0.0027
3	0.2656	0.1736	-0.2510	0.1917	0.0497	-0.0176	0.1446	-0.0212
4	0.0913	0.2579	-0.2668	0.1891	0.1212	-0.0784	0.1987	-0.0031
5	0.1437	0.1220	-0.3169	0.2173	0.1464	-0.0789	0.2190	0.0394
6	0.1761	0.1336	-0.2332	0.1702	0.0216	-0.1478	0.1596	-0.1034
7	0.0520	0.0593	-0.0811	0.1230	0.1334	-0.0629	0.1452	0.1179
8	0.1099	0.0981	-0.3567	0.3214	0.3029	-0.0534	0.2961	-0.0262
9	0.1900	0.0601	-0.2076	0.2486	0.1513	-0.0914	0.2765	-0.0604
10	0.0855	0.1066	-0.2784	0.1898	-0.0231	-0.0866	0.1474	-0.0505
11	0.2154	0.0515	-0.1793	0.1985	0.0490	-0.0690	0.1856	-0.1102
12	0.1567	0.1009	-0.1231	0.0883	0.0743	-0.0542	0.0869	-0.0366
13	0.0976	0.0658	-0.1656	0.1738	0.1873	-0.0324	0.1981	0.0456
14	0.1172	0.0979	-0.2154	0.1918	0.2025	-0.0639	0.2538	0.0972
15	0.1875	0.2016	-0.3495	0.2841	0.1341	-0.0620	0.3185	0.0067
16	0.1370	0.2451	-0.2708	0.1720	0.1494	-0.0880	0.2355	0.0266
17	0.1292	0.1801	-0.1344	0.1946	0.2358	-0.1032	0.2250	0.0220
18	0.1909	0.2071	-0.2565	0.1989	0.1634	-0.1225	0.2446	0.0583
19	0.0748	0.1222	-0.1689	0.1790	0.1661	0.0046	0.1987	-0.0833

TABLE D-3 (CONT.)

PREDICTORS 25-29

NOS	COMB	FSER	PSER	TECH	MACH
1	0.1850	-0.2557	-0.0926	0.0542	-0.1439
2	0.1654	-0.1917	-0.1648	-0.0114	-0.0936
3	0.2440	-0.1549	-0.0493	-0.0094	0.0528
4	0.1297	-0.1199	-0.0496	0.0668	-0.1336
5	0.2227	-0.1647	-0.0257	0.1130	-0.0484
6	0.1615	-0.2151	-0.2447	-0.0292	-0.0948
7	0.1080	-0.0353	0.0471	0.1494	0.0729
8	0.2771	-0.1456	-0.2356	0.0840	-0.0162
9	0.1226	-0.1519	-0.1569	0.0065	-0.0654
10	0.1853	-0.1257	-0.1254	-0.0740	-0.0022
11	0.3035	-0.1900	-0.0830	-0.0954	0.2673
12	0.1551	-0.1512	-0.0357	-0.0283	0.1084
13	0.1159	-0.0768	0.0643	0.0811	-0.0405
14	0.0180	-0.1184	-0.0396	0.1456	-0.1873
15	0.1575	-0.2026	-0.1461	0.0837	-0.0484
16	-0.0168	-0.1834	-0.2037	0.0817	-0.1650
17	0.1873	-0.0870	-0.0321	0.0313	-0.0122
18	-0.0347	-0.0094	-0.1154	0.1022	-0.2367
19	0.0169	-0.1321	-0.0308	-0.0117	-0.1340

APPENDIX E

ANALYSIS SAMPLE GENERATION AND DATA

I. PROCEDURE FOR GENERATING ANALYSIS SAMPLE

a. Notation

N = number of entities in an MOS sample
n = number of test variables (j=1...29)
m = number of jobs (i=1...19)
R_t = 29x29 matrix of predictor intercorrelations
V = m x 29 matrix of validity coefficients
X_m = N x 30 matrix of random normal deviates for the mth job sample

b. **Compute for each of the 19 job samples, 29 synthetic test scores and 1 criterion score using the Gramian factor solution of Rtv_i as the transformation matrix.**

b.1 $Fi = Rtv_i\ (A\ D^{-1/2}\ A')$

where,

Fi = 30x30 transformation matrix for the ith job sample

Rtv_i = matrix of 29 test intercorrelations in rows and columns 1-29 plus vectors of 29 validities in 30th row and column for the ith job;

$$\left[\begin{array}{c|c} R_t & vi \\ \hline vi & 1.0 \end{array}\right]$$

A = eigenvectors of 30x30 intercorrelation-validity matrix

D = diagonal matrix of eigenvalues of intercorrelation-validity matrix

b.2 $Yi = X_N\ Fi'$

where,

Yi = Nx30 matrix of test and criterion scores (with the same expected parameters as the population) for N entities in the ith job sample.

c. **Compute matrix of sums-of-squares and cross-products, $Q_i=Y_i.Y_i$, for each job sample.**

d. **Compute the vector of covariances, Ci, for each job sample.**

d.1 Identify q_i, the 30th row of Qi.

d.2 $m_i = (I'Yi)1/N$

where,

m_i = row vector of means of 29 predictors and

1 criteria (1x30)

1 = summing vector of 1s

d.3 $c_i = (1/N)q_i - (mik\ m_i)$; and drop the 30th element.

where,

m_{ik} = a scalar which is the 30th element of m_i.

c_i = 1x29 vector of covariances of the predictors

e. Compute analysis sample validity matrix (Va) using 19 Cis.

e.1 For each job sample, compute validities for 29 predictors: $v_{ai} = (Si^{-1/2})\ c_i\ (1/s_i)$

where,

v_{ai} = 1x29 vector of validities between 29 predictors and 1 criterion variable for each job sample

Si = diagonal matrix (taken from Qi) with the variances of the 29 predictors and 1 job sample criterion in the diagonal

s_i = scalar which is the covariance of the criterion for the ith job sample; it is the 30th element of the matrix Si

e.2 Assemble 19x29 validity matrix (Va) for combined job samples using v_{ai}.

$Va = (v_{a1'}, v_{a2'}, \ldots, v_{a19'})'$

f. Compute analysis sample intercorrelation matrix (R_a)

f.1 Drop criterion variable from 19 Qis.

f.2 Sum the 19 Qis weighted by sample size:
$Qt = \Sigma_{i=119} Ni\ Qi$

f.3 Compute analysis sample intercorrelation matrix:

a. $m_t = [\Sigma_{i=119} Ni(m_i)]\ 1/\Sigma_{i=119} Ni$

b. $Ct = Qt\ (1/\Sigma_{i=119} Ni) - m_t'(m_t)$

where,

Ct = combined covariance matrix for the 19 job samples

Qt = combined sums-of-squares and cross-products matrix for the 19 job samples

c. $R_a = S^{-1/2}\ Ct\ S^{-1/2}$

where,

S = diagonal matrix of the diagonal elements of Ct (i.e., predictor variances).

TABLE E-1: SEEDS USED TO GENERATE MATRIX OF RANDOM NORMAL DEVIATES (X_m) FOR ANALYSIS SAMPLE

X_m for mth MOS sample	Seed
X1	4230175544 - INITIAL SEED
X2	1430478015
X3	374314139
X4	823870838
X5	1177510215
X6	379516625
X7	1042803379
X8	911214433
X9	2071236145
X10	149236351
X11	613354435
X12	1210466674
X13	1144093513
X14	1732279920
X15	318522932
X16	1220158268
X17	1583150165
X18	1342521610
X19	1463079165

TABLE E-2: ANALYSIS SAMPLE PREDICTOR INTERCORRELATIONS (see Appendix B for code names)

PREDICTORS 1-9

	GS	AR	NO	CS	AS	MK	MC	EI	VE
GS	1.0000	0.7200	0.5200	0.4500	0.6400	0.6900	0.7000	0.7600	0.8000
AR	0.7200	1.0000	0.6300	0.5100	0.5300	0.8300	0.6900	0.6600	0.7300
NO	0.5200	0.6300	1.0000	0.7000	0.3000	0.6200	0.4000	0.4100	0.6200
CS	0.4500	0.5100	0.7000	1.0000	0.2200	0.5200	0.3400	0.3400	0.5700
AS	0.6400	0.5300	0.3000	0.2200	1.0000	0.4100	0.7400	0.7500	0.5200
MK	0.6900	0.8300	0.6200	0.5200	0.4100	1.0000	0.6000	0.5900	0.7000
MC	0.7000	0.6900	0.4000	0.3400	0.7400	0.6000	1.0000	0.7400	0.6000
EI	0.7600	0.6600	0.4100	0.3400	0.7500	0.5900	0.7400	1.0000	0.6700
VE	0.8000	0.7300	0.6200	0.5700	0.5200	0.7000	0.6000	0.6700	1.0000
SPAT	0.6707	0.7301	0.5162	0.4877	0.5677	0.6802	0.7413	0.6159	0.6234
CPAC	0.3166	0.3560	0.3047	0.3155	0.2084	0.3485	0.2775	0.2749	0.3678
CPSP	0.3170	0.2876	0.3119	0.2953	0.2427	0.2811	0.3005	0.2617	0.2783
NMSA	0.5895	0.7156	0.6966	0.5545	0.3938	0.6774	0.4914	0.4996	0.6498
PSYM	0.4459	0.4383	0.3249	0.2920	0.4586	0.3841	0.5479	0.4544	0.3773
SRAC	0.2136	0.2179	0.1653	0.1703	0.1901	0.1861	0.2100	0.2023	0.2312
SRSP	0.2169	0.2283	0.2646	0.2534	0.1385	0.2146	0.1892	0.1766	0.2288
AUTO	0.2486	0.2261	0.1849	0.1562	0.2227	0.1953	0.2203	0.2393	0.2602
SUPP	0.1383	0.1196	0.1739	0.1745	0.0436	0.1294	0.0584	0.0938	0.2090
ROUT	-0.3150	-0.3021	-0.2525	-0.2355	-0.2507	-0.2620	-0.2898	-0.2737	-0.3420
ADJU	0.2256	0.2399	0.1925	0.1338	0.2048	0.2147	0.2261	0.2259	0.2315
DEPN	0.0522	0.1017	0.1350	0.1520	-0.0384	0.1450	0.0162	0.0330	0.0889
COND	-0.0462	-0.0322	-0.0048	-0.0387	-0.0147	-0.0269	-0.0123	-0.0348	-0.0556
SURG	0.2076	0.2533	0.2371	0.2020	0.1593	0.2393	0.1903	0.2003	0.2392
AUDI	0.0147	0.0022	0.0058	0.0221	-0.0909	0.0482	-0.0184	-0.0171	0.0507
COMB	0.1539	0.0660	-0.0309	-0.0663	0.3433	0.0120	0.2594	0.2220	0.0435
FSER	-0.2097	-0.1852	-0.1295	-0.1199	-0.2366	-0.1408	-0.2317	-0.2179	-0.1939
PSER	-0.0990	-0.1426	-0.1365	-0.1275	0.0101	-0.1601	-0.0577	-0.0580	-0.1356
TECH	-0.0039	0.0629	0.1116	0.0783	-0.1353	0.1275	-0.0575	-0.0342	0.0483
MACH	-0.1545	-0.1908	-0.2822	-0.2951	0.1864	-0.2210	0.0465	0.0075	-0.2955

TABLE E-2 (CONT.): ANALYSIS PREDICTOR CORRELATIONS

PREDICTORS 10-18

	SPAT	CPAC	CPSP	NMSA	PSYM	SRAC	SRSP	AUTO	SUPP
GS	0.6707	0.3166	0.3170	0.5895	0.4459	0.2136	0.2169	0.2486	0.1383
AR	0.7301	0.3560	0.2876	0.7156	0.4383	0.2179	0.2283	0.2261	0.1196
NO	0.5162	0.3047	0.3119	0.6966	0.3249	0.1653	0.2646	0.1849	0.1739
CS	0.4877	0.3155	0.2953	0.5545	0.2920	0.1703	0.2534	0.1562	0.1745
AS	0.5677	0.2084	0.2427	0.3938	0.4586	0.1901	0.1385	0.2227	0.0436
MK	0.6802	0.3485	0.2811	0.6774	0.3841	0.1861	0.2146	0.1953	0.1294
MC	0.7413	0.2775	0.3005	0.4914	0.5479	0.2100	0.1892	0.2203	0.0584
EI	0.6159	0.2749	0.2617	0.4996	0.4544	0.2023	0.1766	0.2393	0.0938
VE	0.6234	0.3678	0.2783	0.6498	0.3773	0.2312	0.2288	0.2602	0.2090
SPAT	1.0000	0.3886	0.4057	0.6143	0.6040	0.2311	0.2569	0.2039	0.0978
CPAC	0.3886	1.0000	-0.2025	0.3000	0.2477	0.2284	0.0695	0.0566	0.0886
CPSP	0.4057	-0.2025	1.0000	0.4129	0.3768	0.0642	0.3716	0.1009	0.0516
NMSA	0.6143	0.3000	0.4129	1.0000	0.4413	0.1983	0.3023	0.1890	0.1497
PSYM	0.6040	0.2477	0.3768	0.4413	1.0000	0.1434	0.2696	0.1371	0.0528
SRAC	0.2311	0.2284	0.0642	0.1983	0.1434	1.0000	0.1200	0.0368	0.0561
SRSP	0.2569	0.0695	0.3716	0.3023	0.2696	0.1200	1.0000	0.0681	0.0635
AUTO	0.2039	0.0566	0.1009	0.1890	0.1371	0.0368	0.0681	1.0000	0.2877
SUPP	0.0978	0.0886	0.0516	0.1497	0.0528	0.0561	0.0635	0.2677	1.0000
ROUT	-0.2974	-0.1429	-0.1408	-0.2577	-0.2091	-0.0919	-0.1225	-0.1530	-0.2384
ADJU	0.2258	0.1227	0.1186	0.2093	0.1934	0.0690	0.1149	0.1069	0.1163
DEPN	0.0561	0.1070	-0.0034	0.0990	-0.0230	0.0245	0.0250	0.0051	0.2542
COND	-0.0352	-0.0547	0.0688	0.0062	0.0990	-0.0456	0.0477	0.0531	0.0584
SURG	0.2023	0.1246	0.0997	0.2301	0.1352	0.0498	0.0991	0.2010	0.3358
AUDI	0.0032	0.0199	0.0052	-0.0293	-0.0224	-0.0241	-0.0152	0.1033	0.1682
COMB	0.1737	0.0135	0.0728	0.0276	0.2522	0.0092	0.0196	0.1373	0.0294
FSER	-0.2148	-0.0924	-0.1278	-0.1650	-0.2282	-0.0718	-0 1014	-0.0944	-0.0527
PSER	-0.0907	-0.0818	0.0008	-0.1145	0.0214	-0.0293	-0.0071	0.0002	0.0635
TECH	-0.0134	0.0601	-0.0156	0.0688	-0.0316	-0.0292	0.0116	0.0684	0.2415
MACH	-0.0620	-0.1171	-0.0538	-0.2163	0.0616	-0.065	-0.0809	0.0138	-0.0697

TABLE E-2 (CONT.)

PREDICTORS 19-24

	ROUT	ADJU	DEPN	COND	SURG	AUSI
GS	-0.3150	0.2256	0.0522	-0.0462	0.2076	0.0147
AR	-0.3021	0.2399	0.1017	-0.0322	0.2533	0.0022
NO	-0.2525	0.1925	0.1350	-0.0048	0.2371	0.0058
CS	-0.2355	0.1338	0.1520	-0.0387	0.2020	0.0221
AS	-0.2507	0.2048	-0.0384	-0.0147	0.1593	-0.0909
MK	-0.2620	0.2147	0.1450	-0.0269	0.2393	0.0482
MC	-0.2898	0.2261	0.0162	-0.0123	0.1903	-0.0184
EI	-0.2737	0.2259	0.0330	-0.0348	0.2003	-0.0171
VE	-0.3420	0.2315	0.0889	-0.0556	0.2392	0.0507
SPAT	-0.2974	0.2258	0.0561	-0.0352	0.2023	0.0032
CPAC	-0.1429	0.1227	0.1070	-0.0547	0.1246	0.0199
CPSP	-0.1408	0.1186	-0.0034	0.0688	0.0997	0.0052
NMSA	-0.2577	0.2093	0.0990	0.0062	0.2301	-0.0293
PSYM	-0.2091	0.1934	-0.0230	0.0990	0.1352	-0.0224
SRAC	-0.0919	0.0694	0.0245	-0.0456	0.0498	-0.0241
SRSP	-0.1225	0.1149	0.0250	0.0477	0.0991	-0.0152
AUTO	-0.1530	0.1069	0.0051	0.0531	0.2010	0.1033
SUPP	-0.2384	0.1163	0.2542	0.0584	0.3358	0.1682
ROUT	1.0000	-0.1912	-0.0363	-0.0653	-0.2435	-0.0059
ADJU	-0.1912	1.0000	0.3414	0.2268	0.6038	0.0622
DEPN	-0.0363	0.3414	1.0000	0.1279	0.5971	0.1924
COND	-0.0653	0.2268	0.1279	1.0000	0.3410	0.0622
SURG	-0.2435	0.6038	0.5971	0.3410	1.0000	0.1838
AUDI	-0.0059	0.0622	0.1924	0.0622	0.1838	1.0000
COMB	-0.0808	0.1666	-0.0298	0.1537	0.1868	0.1781
FSER	0.2245	-0.0707	0.0489	-0.0341	-0.0412	0.3074
PSER	0.0482	0.0392	0.0340	0.1304	0.0790	0.1378
TECH	0.0084	0.1489	0.3069	0.0869	0.2955	0.6719
MACH	0.1119	0.0014	-0.1022	0.1296	0.0076	0.2014

TABLE E-2 (CONT.): ANALYSIS PREDICTOR CORRELATIONS

PREDICTORS 25-29

	COMB	FSER	PSER	TECH	MACH
GS	0.1539	-0.2097	-0.0990	-0.0039	-0.1545
AR	0.0660	-0.1852	-0.1426	0.0629	-0.1908
MO	-0.0309	-0.1295	-0.1365	0.1116	-0.2822
CS	-0.0663	-0.1199	-0.1275	0.0783	-0.2951
AS	0.3433	-0.2366	0.0101	-0.1353	0.1864
MK	0.0120	-0.1408	-0.1601	0.1275	-0.2210
MC	0.2594	-0.2317	-0.0577	-0.0575	0.0465
EI	0.2220	-0.2179	-0.0580	-0.0342	0.0075
VE	0.0435	-0.1939	-0.1356	0.0483	-0.2955
SPAT	0.1737	-0.2148	-0.0907	-0.0134	-0.0620
CPAC	0.0135	-0.0924	-0.0818	0.0601	-0.1171
CPSP	0.0728	-0.1278	0.0008	-0.0156	-0.0538
MMSA	0.0276	-0.1650	-0.1145	0.0688	-0.2163
PSYM	0.2522	-0.2282	0.0214	-0.0316	0.0616
SRAC	0.0092	-0.0718	-0.0293	-0.0292	-0.0653
SRSP	0.0196	-0.1014	-0.0071	0.0116	-0.0809
AUTO	0.1373	-0.0944	0.0002	0.0684	0.0138
SUPP	0.0294	-0.0527	0.0635	0.2415	-0.0697
ROUT	-0.0808	0.2245	0.0482	0.0084	0.1119
ADJU	0.1666	-0.0707	0.0392	0.1489	0.0014
DEPN	-0.0298	0.0489	0.0340	0.3069	-0.1022
COND	0.1537	-0.0341	0.1304	0.0869	0.1296
SURG	0.1868	-0.0412	0.0790	0.2955	0.0076
AUDI	0.1781	0.3074	0.1378	0.6719	0.2014
COMB	1.0000	0.0864	0.3913	0.1905	0.5881
FSER	0.0864	1.0000	0.1708	0.3518	0.2269
PSER	0.3913	0.1708	1.0000	0.2216	0.3364
TECH	0.1905	0.3518	0.2216	1.0000	0.2118
MACH	0.5881	0.2269	0.3364	0.2118	1.0000

TABLE E-3: ANALYSIS SAMPLE VALIDITY COEFFICIENTS FOR 19 MOS USING THE SINGLE (CTP) CRITERION (see Appendices A & B for code names)

PREDICTORS 1-9

MOS	GS	AR	MO	CS	AS	MK	MC	EI	VE
1	0.6474	0.6285	0.5424	0.4772	0.4808	0.6309	0.5561	0.5711	0.6361
2	0.6895	0.6255	0.4760	0.3488	0.5842	0.6285	0.6265	0.6536	0.6584
3	0.4179	0.3814	0.3386	0.2657	0.4336	0.3251	0.3897	0.3833	0.4072
4	0.4960	0.5481	0.4003	0.3978	0.3307	0.5517	0.4020	0.3993	0.4864
5	0.6123	0.5318	0.4251	0.3451	0.4856	0.5476	0.5301	0.5492	0.5398
6	0.6587	0.6195	0.6300	0.5674	0.5174	0.5688	0.5604	0.6152	0.6547
7	0.4723	0.5421	0.3194	0.1876	0.3727	0.5271	0.4452	0.4761	0.4437
8	0.7057	0.6665	0.7260	0.6146	0.5683	0.6565	0.6588	0.5572	0.7119
9	0.5859	0.6141	0.4297	0.3863	0.4947	0.5936	0.5133	0.5603	0.5425
10	0.4819	0.4067	0.3632	0.3585	0.3731	0.3643	0.4531	0.4325	0.5438
11	0.4660	0.4143	0.2207	0.2500	0.5822	0.3545	0.5521	0.5134	0.3755
12	0.2972	0.3480	0.0885	0.0819	0.3786	0.3209	0.4033	0.3495	0.2284
13	0.3958	0.3876	0.2425	0.2096	0.3495	0.3895	0.4042	0.4161	0.3595
14	0.4517	0.5286	0.3734	0.3276	0.2269	0.5674	0.3453	0.3253	0.4752
15	0.7178	0.6852	0.4269	0.4707	0.7020	0.6392	0.7028	0.6818	0.6940
16	0.5710	0.6108	0.4197	0.4012	0.4416	0.6335	0.4854	0.5559	0.5772
17	0.4682	0.4395	0.3757	0.4378	0.3897	0.4701	0.4124	0.4189	0.4361
18	0.5546	0.6800	0.5202	0.5034	0.3976	0.6157	0.4891	0.4755	0.6117
19	0.3057	0.3562	0.3345	0.2795	0.2316	0.3607	0.2710	0.3117	0.3283

TABLE E-3 (CONT.): ANALYSIS VALIDITY COEFFICIENTS

PREDICTORS 10-18

MOS	SPAT	CPAC	CPSP	NHSA	PSYM	SRAC	SRSP	AUTO	SUPP
1	0.6646	0.3630	0.3415	0.5357	0.4177	0.2229	0.2483	0.2029	0.1417
2	0.6217	0.3127	0.2198	0.5008	0.3574	0.1782	0.1816	0.2236	0.1086
3	0.4817	0.2669	0.2163	0.3855	0.3377	0.1300	0.2321	0.2656	0.1736
4	0.5444	0.3132	0.1611	0.4858	0.4078	0.0871	0.1181	0.0913	0.2579
5	0.5841	0.3791	0.2182	0.5249	0.3569	0.2296	0.1966	0.1437	0.1220
6	0.5230	0.3041	0.2643	0.5701	0.4099	0.1813	0.1580	0.1761	0.1336
7	0.4493	0.2642	0.1076	0.4203	0.2766	0.0926	0.1154	0.0520	0.0593
8	0.7750	0.5240	0.1402	0.6348	0.4727	0.1465	0.2465	0.1099	0.0981
9	0.5921	0.3127	0.2333	0.5284	0.3685	0.2311	0.1295	0.1900	0.0601
10	0.5080	0.3242	0.1341	0.3653	0.4061	0.2264	0.1407	0.0855	0.1066
11	0.5238	0.1494	0.2356	0.3101	0.3578	0.1643	0.1850	0.2154	0.0515
12	0.4006	0.2394	0.1842	0.2639	0.2504	0.1542	0.0858	0.1567	0.1009
13	0.4551	0.2167	0.0433	0.2940	0.2838	0.0930	0.0473	0.0976	0.0658
14	0.5188	0.3649	0.1403	0.4078	0.2740	0.1775	0.1540	0.1172	0.0979
15	0.7169	0.3085	0.3548	0.6012	0.5001	0.3682	0.1713	0.1875	0.2016
16	0.5423	0.3185	0.1671	0.5631	0.2525	0.2379	0.2133	0.1370	0.2451
17	0.5034	0.2270	0.1912	0.4151	0.3101	0.1525	0.1563	0.1292	0.1801
18	0.6485	0.4664	0.2430	0.6001	0.3161	0.3066	0.2250	0.1909	0.2071
19	0.3737	0.2396	0.1293	0.3444	0.2507	0.1044	0.0985	0.0748	0.1222

TABLE E-3 (CONT.)

PREDICTORS 19-25

MOS	ROUT	ADJU	DEPH	COND	SURG	AUSI	COMB
1	-0.2976	0.2241	0.1675	0.0355	0.2971	0.0138	0.1850
2	-0.3019	0.1965	0.0468	-0.0279	0.2150	0.0027	0.1654
3	-0.2510	0.1917	0.0497	-0.0176	0.1446	-0.0212	0.2440
4	-0.2668	0.1891	0.1212	-0.0784	0.1987	-0.0031	0.1297
5	-0.3169	0.2173	0.1464	-0.0789	0.2190	0.0394	0.2227
6	-0.2332	0.1702	0.0216	-0.1478	0.1596	-0.1034	0.1615
7	-0.0811	0.1230	0.1334	-0.0629	0.1452	0.1179	0.1080
8	-0.3567	0.3214	0.3029	-0.0534	0.2961	-0.0262	0.2771
9	-0.2076	0.2486	0.1513	-0.0914	0.2765	0.0604	0.1226
10	-0.2784	0.1898	-0.0231	-0.0866	0.1474	-0.0505	0.1853
11	-0.1793	0.1985	0.0490	[illegible]	0.1856	-0.1102	0.3035
12	-0.1231	0.0883	0.0743	-0.0542	0.0869	-0.0366	0.1551
13	-0.1656	0.1738	0.1873	-0.0324	0.1981	0.0456	0.1159
14	-0.2154	0.1918	0.2025	-0.0639	0.2538	0.0972	0.0180
15	-0.3495	0.2841	0.1341	-0.0620	0.3185	0.0067	0.1575
16	-0.2708	0.1720	0.1494	-0.0880	0.2355	0.0263	-0.0168
17	-0.1344	0.1946	0.2358	-0.1032	0.2250	0.0220	0.1873
18	-0.2565	0.1989	0.1634	-0.1225	0.2446	0.0583	-0.0347
19	-0.1689	0.1790	0.1661	0.0046	0.1987	-0.0833	0.0169

TABLE E-3 (CONT.): ANALYSIS VALIDITY COEFFICIENTS

MOS	FSER	PSER	TECH	MACH
	PREDICTORS 26-29			
1	-0.2557	-0.0926	0.0542	-0.1439
2	-0.1917	-0.1648	-0.0114	-0.0936
3	-0.1549	-0.0493	-0.0094	0.0528
4	-0.1199	-0.0496	0.0668	-0.1336
5	-0.1647	-0.0257	0.1130	-0.0484
6	-0.2151	-0.2447	-0.0292	-0.0948
7	-0.0353	-0.0471	0.1494	0.0729
8	-0.1456	-0.2356	0.0840	-0.0162
9	-0.1519	-0.1669	0.0065	-0.0654
10	-0.1257	-0.1254	-0.0740	-0.0022
11	-0.1900	-0.0830	-0.0954	0.2673
12	-0.1512	-0.0357	-0.0283	0.1084
13	-0.0768	-0.0643	0.0811	-0.0405
14	-0.1184	-0.0396	0.1456	-0.1873
15	-0.2026	-0.1461	0.0837	-0.0484
16	-0.1834	-0.2037	0.0817	-0.1650
17	-0.0870	-0.0321	0.0313	-0.0122
18	-0.0094	-0.1154	0.1022	-0.2367
19	-0.1321	-0.0308	-0.0117	-0.1340

APPENDIX F

TEST SELECTION FORMULAE AND RESULTS

TABLE F-1 TEST SELECTION INDEX FORMULAE

Index	Formula
Point Distance Index (PDI)	$PDI = \Sigma^{m}_{j=1} (\Sigma^{k}_{i=1} (a_{ij} - a_j*)^2)^{1/2}$ where, a_{ij} = the elements of factor matrix F with i rows equal to the number of jobs and j columns equal to k, the number of selected tests plus the trial test, $a_j* = \Sigma^{m}_{i=1} a_{ij}$, where m equals the number of jobs.
Horst's Index of Differential Efficiency (H_d)	$H_d = \Sigma^{k}_{j=1} \Sigma^{m}_{i=1} (a_{ij} - a_j*)^2$ (defined as above)
Modified Point Distance Index (Mod. PDI)	$\text{Mod. PDI} = \Sigma^{m}_{j=1} (\Sigma^{k}_{i=1} ((R^*/R_i) a_{ij} - a_j*)^2)^{1/2}$ where, $R_i = (\Sigma^{k}_{j=1} (a_{ij})^2)^{1/2}$, $R^* = \Sigma^{m}_{i=1} R_i$.
Modified Horst's Differential Index (Mod. H_d)	$\text{Mod. } H_d = \Sigma^{k}_{j=1} \Sigma^{m}_{i=1} (aij(R*/Ri) - a_j*)^2$ (defined as above)
Max-PSE	$\text{Max-PSE} = 1/m \; \Sigma^{m}_{i=1} (\Sigma^{k}_{j=1} (a_{ij})^2)^{1/2}$ (defined as above)

I. TEST SELECTION RESULTS: COMPOSITE CRITERION VERSUS CORE TECHNICAL PROFICIENCY

1. Tests and Test Order[1]

TABLE F-2: ORDER OF TESTS SELECTED - COMPOSITE CRITERION

Tests Selected		PDI	Hd	Mod. PDI	Mod. Hd	Max PSE
GS	General Science	5	8		8	
AR	Arithmetic Reasoning	6	7		7	1
NO	Number Operations	1	1	7	5	6
CS	Coding Speed	3	4	5	3	10
AS	Auto Shop					5
MK	Math Knowledge	2	2	6	4	7
SPAT	Spatial composite					3
CPSP	Complex perceptual speed				10	
SRAC	Simple reaction accuracy			4		
ROUT	Routine composite	7	10	8	6	8
ADJU	Adjustment composite				9	
DEPN	Dependability composite			10		4
COND	Physical condition			3		
SURG	Achievement orientation					2
COMB	Combat interest	10	6			9
FSER	Food service interest	9	9	9	2	
PSER	Protective service interest	8	5	2		
MACH	Machinery interest	4	3	1	1	

TABLE F-3: ORDER OF TESTS SELECTED - SINGLE CRITERION (CTP)

Tests Selected		PDI	Hd	Mod. PDI	Mod. Hd	Max. PSE
AR	Arithmetic Reasoning					1
NO	Number Operations	1	1	8	2	6
CS	Coding Speed	4	5	10	6	
AS	Auto Shop	2	2	7	4	4
MK	Math Knowledge	3	3	9	3	7
VE	Verbal	8	8		5	3
SPAT	Spatial composite				8	2
CPAC	Complex perceptual accuracy					10
CPSP	Complex perceptual speed		10		9	
SUPP	Support composite	10		6		
DEPN	Dependability composite			2		5
COMB	Combat interest	5	7	4	10	
PSER	Protective service interest	7	6	3		9
TECH	Technical interest	9	9	5	7	
MACH	Machinery interest	6	4	1	1	8

[1] The test selection results in this section were obtained from the corrected Project A data (the universe estimate in this study). (See Appendix E).

2. Factor Solutions for "Best" Ten Tests - Composite

PDI	NO	MK	CS	MACH	GS	AR	ROUT	PSER	FSER	COMB
NO	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
MK	.620	.785	.000	.000	.000	.000	.000	.000	.000	.000
CS	.700	.110	.706	.000	.000	.000	.000	.000	.000	.000
MACH	-.282	-.059	-.129	.949	.000	.000	.000	.000	.000	.000
GS	.520	.469	.049	.027	.712	.000	.000	.000	.000	.000
AR	.630	.560	.011	.022	.181	.506	.000	.000	.000	.000
ROUT	-.252	-.134	-.062	.026	-.166	-.074	.938	.000	.000	.000
PSER	-.137	-.096	-.030	.304	.014	-.023	-.009	.937	.000	.000
FSER	-.130	-.077	-.029	.192	-.155	-.072	.153	.094	.937	.000
COMB	-.031	.040	-.069	.604	.194	.030	-.073	.216	-.010	.733

HD	NO	MK	MACH	CS	PSER	COMB	AR	GS	FSER	ROUT
NO	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
MK	.620	.785	.000	.000	.000	.000	.000	.000	.000	.000
MACH	-.282	-.059	.958	.000	.000	.000	.000	.000	.000	.000
CS	.700	.110	-.095	.699	.000	.000	.000	.000	.000	.000
PSER	-.137	-.096	.305	.011	.937	.000	.000	.000	.000	.000
COMB	-.031	.040	.607	.013	.219	.762	.000	.000	.000	.000
AR	.630	.560	.021	.014	-.010	.069	.533	.000	.000	.000
GS	.520	.469	.021	.052	.011	.178	.219	.654	.000	.000
FSER	-.130	-.077	.194	-.003	.092	-.069	-.110	-.114	.949	.000
ROUT	-.252	-.134	.034	-.058	-.010	-.133	-.110	-.108	.149	.922

M.PDI	MACH	PSER	COND	SRAC	CS	MK	NO	ROUT	FSER	DEPN
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PSER	.336	.942	.000	.000	.000	.000	.000	.000	.000	.000
COND	.130	.092	.987	.000	.000	.000	.000	.000	.000	.000
SRAC	-.065	-.008	-.037	.997	.000	.000	.000	.000	.000	.000
CS	-.295	-.030	.002	.151	.943	.000	.000	.000	.000	.000
MK	-.221	-.091	.010	.172	.452	.842	.000	.000	.000	.000
NO	-.282	-.044	.036	.148	.629	.289	.645	.000	.000	.000
ROUT	.112	.011	-.082	-.088	-.200	-.154	-.053	.952	.000	.000
FSER	.227	.100	-.074	-.059	-.043	-.061	-.008	.177	.945	.000
DEPN	-.102	.073	.136	.023	.127	.078	-.003	.026	.087	.967

M.HD	MACH	FSER	CS	MK	NO	ROUT	AR	GS	ADJU	CPSP
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
FSER	.227	.974	.000	.000	.000	.000	.000	.000	.000	.000
CS	-.295	-.054	.954	.000	.000	.000	.000	.000	.000	.000
MK	-.221	-.093	.471	.849	.000	.000	.000	.000	.000	.000
NO	-.282	-.067	.643	.293	.546	.000	.000	.000	.000	.000
ROUT	.112	.204	-.201	-.146	-.055	.939	.000	.000	.000	.000
AR	-.191	-.146	.467	.653	.116	-.059	.530	.000	.000	.000
GS	-.154	-.179	.414	.523	.070	-.104	.217	.660	.000	.000
ADJU	.001	-.073	.137	.170	.078	-.128	.072	.050	958	.000
CPSP	-.054	-.119	.286	.145	.096	-.028	.035	.115	.028	.925

M-PSE	AR	SURG	SPAT	DEPN	AS	NO	MK	ROUT	COMB	CS
AR	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
SURG	.253	.967	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.730	.018	.683	.000	.000	.000	.000	.000	.000	.000
DEPN	.102	.591	-.042	.799	.000	.000	.000	.000	.000	.000
AS	.530	.026	.264	-.121	.796	.000	.000	.000	.000	.000
NO	.630	.080	.080	.034	-.067	.765	.000	.000	.000	.000
MK	.830	.030	.108	.059	-.065	.104	.529	.000	.000	.000
ROUT	-.302	-.173	-.108	.115	-.055	-.062	.003	.920	.000	.000
COMB	.066	.176	.179	-.166	.297	-.099	-.053	.020	.897	.000
CS	.510	.075	.167	.078	-.109	.457	.032	-.041	-.056	.688

3. Factor Solutions for "Best" Ten Tests - CTP

PDI	NO	AS	MK	CS	COMB	MACH	PSER	VE	TECH	SUPP
NO	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
AS	.300	.954	.000	.000	.000	.000	.000	.000	.000	.000
MK	.620	.235	.749	.000	.000	.000	.000	.000	.000	.000
CS	.700	.010	.112	.705	.000	.000	.000	.000	.000	.000
COMB	-.031	.370	-.074	-.057	.924	.000	.000	.000	.000	.000
MACH	-.282	.284	-.151	-.119	.494	.748	.000	.000	.000	.000
PSER	-.137	.054	-.118	-.027	.386	.095	.898	.000	.000	.000
VE	.620	.350	.312	.138	-.039	-.184	.003	.584	.000	.000
TECH	.112	-.177	.133	-.018	.290	.225	.143	.093	.880	.000
SUPP	.174	-.009	.032	.070	.048	-.038	.087	.136	.213	.943

HD	NO	AS	MK	MACH	CS	PSER	COMB	VE	TECH	CPSP
NO	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
AS	.300	.954	.000	.000	.000	.000	.000	.000	.000	.000
MK	.620	.235	.749	.000	.000	.000	.000	.000	.000	.000
MACH	-.282	.284	-.151	.904	.000	.000	.000	.000	.000	.000
CS	.700	.010	.112	-.093	.699	.000	.000	.000	.000	.000
PSER	-.137	.054	-.118	.293	.011	.937	.000	.000	.000	.000
COMB	-.031	.370	-.074	.512	.010	.222	.738	.000	.000	.000
VE	.620	.350	.312	-.191	.114	.023	.065	.584	.000	.000
TECH	.112	-.177	.133	.347	.028	.171	.072	.093	.880	.000
CPSP	.312	.156	.068	.000	.097	.045	.025	-.008	-.049	.927

M.PDI	MACH	DEPN	PSER	COMB	TECH	SUPP	AS	NO	MK	CS
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
DEPN	-.102	.995	.000	.000	.000	.000	.000	.000	.000	.000
PSER	.336	.069	.939	.000	.000	.000	.000	.000	.000	.000
COMB	.588	.030	.204	.782	.000	.000	.000	.000	.000	.000
TECH	.212	.330	.136	.036	.909	.000	.000	.000	.000	.000
SUPP	-.070	.248	.074	.061	.178	.945	.000	.000	.000	.000
AS	.186	-.019	-.055	.314	-.189	.085	.906	.000	.000	.000
NO	-.282	.107	-.052	.182	.150	.099	.348	.848	.000	.000
MK	-.221	.123	-.100	.203	.154	.054	.452	.373	.717	.000
CS	-.295	.122	-.039	.143	.111	.104	.268	.537	.087	.697

M.HD	MACH	NO	MK	AS	VE	CS	TECH	SPAT	CPSP	COMB
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
NO	-.282	.959	.000	.000	.000	.000	.000	.000	.000	.000
MK	-.221	.581	.783	.000	.000	.000	.000	.000	.000	.000
AS	.186	.368	.303	.859	.000	.000	.000	.000	.000	.000
VE	-.296	.559	.395	.291	.599	.000	.000	.000	.000	.000
CS	-.295	.643	.104	.009	.133	.686	.000	.000	.000	.000
TECH	.212	.179	.090	-.312	.110	.007	.898	.000	.000	.000
SPAT	-.062	.520	.465	.288	.078	.108	-.061	.637	.000	.000
CPSP	-.054	.309	.114	.122	.015	.096	-.038	.219	.903	.000
COMB	.588	.141	.077	.185	.091	-.007	.091	.074	.018	.756

M-PSE	AR	SPAT	VE	AS	DEPN	NO	MK	MACH	PSER	CPAC
AR	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.730	.683	.000	.000	.000	.000	.000	.000	.000	.000
VE	.730	.132	.671	.000	.000	.000	.000	.000	.000	.000
AS	.530	.265	.146	.792	.000	.000	.000	.000	.000	.000
DEPN	.102	-.027	.027	-.113	.988	.000	.000	.000	.000	.000
NO	.630	.082	.223	-.111	.055	.729	.000	.000	.000	.000
MK	.830	.109	.119	-.096	.050	.066	.519	.000	.000	.000
MACH	-.191	.113	-.255	.372	-.031	-.098	-.002	.858	.000	.000
PSER	-.143	.020	-.051	.111	.064	-.039	-.054	.292	.933	.000
CPAC	.356	.188	.124	-.061	.066	.037	.012	-.012	-.022	.901

II. TEST SELECTION FACTOR SOLUTIONS BY SELECTION INDEX[2] (see Appendix B for code names)

POINT DISTANCE INDEX

9-JOBS; HANDS-ON TESTING

	MACH	CS	TECH	GS	AR	AUTO	DEPM	MO	COMO	MC
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CS	-.307	.952	.000	.000	.000	.000	.000	.000	.000	.000
TECH	.244	.130	.961	.000	.000	.000	.000	.000	.000	.000
GS	-.161	.425	-.051	.889	.000	.000	.000	.000	.000	.000
AR	-.191	.487	.016	.550	.651	.000	.000	.000	.000	.000
AUTO	.023	.156	.028	.192	.081	.965	.000	.000	.000	.000
DEPM	-.097	.103	.309	-.009	.039	-.016	.939	.000	.000	.000
MO	-.283	.651	.073	.235	.205	.011	-.014	.627	.000	.000
COMO	.131	.001	.030	-.017	.029	.073	.140	.031	.977	.000
MC	.034	.376	-.143	.601	.281	.014	.004	-.021	.004	.630

9-JOBS; NO HANDS-ON TESTING

	AS	MO	PSYM	SPAT	MK	CS	VE	SRAC	PSER	COMB
AS	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
MO	.321	.947	.000	.000	.000	.000	.000	.000	.000	.000
PSYM	.456	.202	.866	.000	.000	.000	.000	.000	.000	.000
SPAT	.573	.370	.302	.667	.000	.000	.000	.000	.000	.000
MK	.424	.522	.110	.323	.657	.000	.000	.000	.000	.000
CS	.230	.668	.074	.157	.046	.685	.000	.000	.000	.000
VE	.528	.484	.052	.197	.242	.126	.609	.000	.000	.000
SRAC	.203	.139	.041	.116	-.002	.048	.065	.958	.000	.000
PSER	-.002	-.129	.064	-.079	-.087	-.054	-.055	.008	.980	.000
COMB	.350	-.144	.144	-.003	-.087	-.083	-.049	-.010	.364	.829

18 JOBS; 9 JOBS WITH, 9 JOBS WITHOUT HANDS-ON TESTING

	CS	AS	VE	MK	PSYM	SPAT	MO	DEPM	COMO	MACH
CS	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
AS	.230	.973	.000	.000	.000	.000	.000	.000	.000	.000
VE	.577	.406	.709	.000	.000	.000	.000	.000	.000	.000
MK	.536	.309	.382	.687	.000	.000	.000	.000	.000	.000
PSYM	.305	.397	.066	.122	.855	.000	.000	.000	.000	.000
SPAT	.506	.469	.206	.276	.246	.587	.000	.000	.000	.000
MO	.706	.163	.218	.173	.027	.002	.630	000	.000	.000
DEPM	.128	-.088	.053	.109	-.053	.001	.008	.979	.000	.000
COMO	-.039	.018	-.057	.017	.145	-.063	.026	.147	.973	.000
MACH	-.307	.256	-.330	-.014	.072	.049	-.056	-.012	.094	.844

[2] The experimental test selection results were obtained from the analysis sample data set given in Appendix F, Tables F-2 & F-3. Results are categorized by the selection index and the job sample to which the index was applied.

HORST'S DIFFERENTIAL INDEX (H_d)

9 JOBS; HANDS-ON TESTING

	MACH	CS	TECH	GS	AR	AUTO	COND	DEPM	NO	MC
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CS	-.307	.952	.000	.000	.000	.000	.000	.000	.000	.000
TECH	.244	.130	.961	.000	.000	.000	.000	.000	.000	.000
GS	-.161	.425	-.051	.889	.000	.000	.000	.000	.000	.000
AR	-.191	.487	.016	.550	.651	.000	.000	.000	.000	.000
AUTO	.023	.156	.028	.192	.081	.965	.000	.000	.000	.000
COND	.131	.001	.030	-.017	.029	.073	.988	.000	.000	.000
DEPM	-.097	.103	.309	-.009	.039	-.016	.133	.930	.000	.000
NO	-.283	.651	.073	.235	.205	.011	.018	-.017	.627	.000
MC	.034	.376	-.143	.601	.281	.014	.004	.004	-.021	.630

9 JOBS; NO HANDS-ON TESTING

	AS	NO	SPAT	PSYM	MK	VE	CS	SRAC	DEPM	GS
AS	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
NO	.321	.947	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.573	.370	.732	.000	.000	.000	.000	.000	.000	.000
PSYM	.456	.202	.357	.790	.000	.000	.000	.000	.000	.000
MK	.424	.522	.340	-.032	.657	.000	.000	.000	.000	.000
VE	.528	.484	.201	-.034	.242	.622	.000	.000	.000	.000
CS	.230	.668	.174	.003	.046	.138	.671	.000	.000	.000
SRAC	.203	.139	.122	-.011	-.002	.073	.033	.958	.000	.000
DEPM	-.056	.135	.047	-.054	.110	.004	.056	.008	.979	.000
GS	.639	.340	.247	.006	.254	.298	-.024	.010	-.013	.512

18 JOBS; 9 JOBS WITH, 9 JOBS WITHOUT HANDS-ON TESTING

	CS	AS	VE	SPAT	MK	PSYM	NO	GS	DEPM	COND
CS	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
AS	.230	.973	.000	.000	.000	.000	.000	000	.000	.000
VE	.577	.406	.709	.000	.000	.000	.000	.000	.000	.000
SPAT	.506	.469	.206	.694	.000	.000	.000	.000	.000	.000
MK	.536	.309	.382	.273	.630	.000	.000	.000	.000	.000
PSYM	.305	.397	.066	.351	-.019	.789	.000	.000	.000	.000
NO	.706	.163	.218	.080	.154	.024	.630	.000	.000	.000
GS	.454	.549	.442	.135	.127	.028	-.016	.512	.000	.000
DEPM	.128	-.088	.053	.026	.108	-.049	.008	-.025	.979	.000
COND	-.039	.018	-.057	.004	.016	.158	.026	-.022	.146	.973

MODIFIED POINT DISTANCE INDEX

9 JOBS; HANDS-ON TESTING

	MACH	TECH	COMS	AUTO	DEPM	COMD	CS	AR	GS	MO
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
TECH	.244	.970	.000	.000	.000	.000	.000	.000	.000	.000
COMS	.574	.065	.816	.000	.000	.000	.000	.000	.000	.000
AUTO	.023	.049	.168	.984	.000	.000	.000	.000	.000	.000
DEPM	-.097	.320	.005	-.005	.942	.000	.000	.000	.000	.000
COMD	.131	.030	.093	.055	.139	.975	.000	.000	.000	.000
CS	-.307	.128	.124	.124	.061	-.030	.924	.000	.000	.000
AR	-.191	.082	.232	.195	.053	-.032	.428	.823	.000	.000
GS	-.161	.007	.328	.183	.019	-.061	.365	.515	.657	.000
MO	-.283	.160	.156	.144	.033	-.009	.605	.294	.045	.627

9 JOBS; NO HANDS-ON TESTING

	DEPM	PSER	SRAC	COMD	MO	MK	PSYM	SPAT	AS	CS
DEPM	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PSER	.026	1.000	.000	.000	.000	.000	.000	.000	.000	.000
SRAC	.024	-.023	.999	.000	.000	.000	.000	.000	.000	.000
COMD	.128	.132	-.012	.983	.000	.000	.000	.000	.000	.000
MO	.110	-.126	.191	-.009	.967	.000	.000	.000	.000	.000
MK	.137	-.147	.192	-.022	.579	.766	.000	.000	.000	.000
PSYM	-.024	.029	.157	.121	.326	.243	.891	.000	.000	.000
SPAT	.052	-.084	.254	-.013	.485	.437	.335	.622	.000	.000
AS	-.056	-.000	.204	.019	.298	.288	.285	.254	.798	.000
CS	.128	-.139	.187	-.036	.660	.102	.052	.092	-.081	.681

18 JOBS; 9 JOBS WITH, 9 JOBS WITHOUT HANDS-ON TESTING

	MACH	DEPM	PSER	TECH	SRAC	MO	AS	PSYM	CS	SPAT
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
DEPM	-.097	.995	.000	.000	.000	.000	.000	.000	.000	.000
PSER	.328	.058	.943	.000	.000	.000	.000	.000	.000	.000
TECH	.244	.312	.153	.905	.000	.000	.000	.000	.000	.000
SRAC	-.051	.019	-.008	-.040	.998	.000	.000	.000	.000	.000
MO	-.283	.083	-.037	.149	.187	.924	.000	.000	.000	.000
AS	.178	-.039	-.061	-.172	.205	.389	.860	.000	.000	.000
PSYM	.046	-.020	.015	-.040	.157	.357	.314	.863	.000	.000
CS	-.307	.099	-.043	.110	.180	.605	.038	.081	.689	.000
SPAT	-.061	.047	-.069	-.028	.252	.506	.381	.303	.134	.644

MODIFIED HORST'S DIFFERENTIAL INDEX

9 JOBS; HANDS-ON TESTING

	MACH	COMB	TECH	AUTO	CS	NO	MC	AR	GS	DEPN
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
COMB	.574	.819	.000	.000	.000	.000	.000	.000	.000	.000
TECH	.244	.077	.967	.000	.000	.000	.000	.000	.000	.000
AUTO	.023	.171	.035	.984	.000	.000	.000	.000	.000	.000
CS	-.307	.134	.118	.124	.927	.000	.000	.000	.000	.000
NO	-.283	.169	.147	.144	.606	.694	.000	.000	.000	.000
MC	.034	.289	-.114	.163	.337	.230	.842	.000	.000	.000
AR	-.191	.237	.063	.195	.431	.348	.445	.599	.000	.000
GS	-.161	.327	-.019	.183	.367	.260	.464	.212	.608	.000
DEPN	-.097	.030	.319	-.005	.062	-.009	.017	.032	-.028	.939

9 JOBS; NO HANDS-ON TESTING

	DEPN	VE	SPAT	MK	NO	PSYM	AS	CS	SRAC	COMD
DEPN	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
VE	.076	.997	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.052	.626	.778	.000	.000	.000	.000	.000	.000	.000
MK	.137	.697	.310	.632	.000	.000	.000	.000	.000	.000
NO	.110	.622	.179	.201	.727	.000	.000	.000	.000	.000
PSYM	-.024	.387	.458	-.021	.031	.799	.000	.000	.000	.000
AS	-.056	.534	.310	-.057	-.067	.134	.768	.000	.000	.000
CS	.128	.569	.184	.102	.392	-.008	-.118	.670	.000	.000
SRAC	.024	.243	.133	-.024	.033	-.000	.044	.033	.958	.000
COMD	.128	-.065	.019	-.005	.014	.171	.030	-.033	-.002	.973

18 JOBS; 9 JOBS WITH, 9 JOBS WITHOUT HANDS-ON TESTING

	MACH	CS	SPAT	MK	GS	NO	PSYM	AS	DEPN	TECH
MACH	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CS	-.307	.952	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	-.061	.512	.857	.000	.000	.000	.000	.000	.000	.000
MK	-.221	.491	.489	.686	.000	.000	.000	.000	.000	.000
GS	-.161	.425	.519	.292	.663	.000	.000	.000	.000	.000
NO	-.283	.651	.214	.209	.049	.636	.000	.000	.000	.000
PSYM	.046	.335	.500	-.007	.091	.036	.791	.000	.000	.000
AS	.178	.299	.502	.103	.376	.046	.078	.683	.000	.000
DEPN	-.097	.103	-.008	.099	-.073	-.000	-.055	-.065	.978	.000
TECH	.244	.130	-.103	.197	-.079	.087	-.034	-.229	.259	.860

MAX-PSE

9 JOBS; HANDS-ON TESTING

	AR	SPAT	EI	MACH	VE	DEPM	TECH	CS	CPAC	AUTO
AR	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.730	.683	.000	.000	.000	.000	.000	.000	.000	.000
EI	.659	.196	.726	.000	.000	.000	.000	.000	.000	.000
MACH	-.191	.115	.154	.963	.000	.000	.000	.000	.000	.000
VE	.733	.136	.207	-.223	.593	.000	.000	.000	.000	.000
DEPM	.095	-.025	-.057	-.070	.010	.991	.000	.000	.000	.000
TECH	.032	-.088	-.072	.282	.126	.298	.895	.000	.000	.000
CS	.522	.183	-.052	-.229	.218	.063	.070	.763	.000	.000
CPAC	.353	.196	-.005	-.054	.126	.062	.048	.078	.898	.000
AUTO	.230	.053	.088	.049	.101	-.004	.035	.005	-.053	.959

9 JOBS; NO HANDS-ON TESTING

	AR	GS	SPAT	CS	VE	PSYM	AS	MK	PSER	DEPM
AR	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
GS	.726	.688	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.730	.206	.651	.000	.000	.000	.000	.000	.000	.000
CS	.522	.109	.157	.831	.000	.000	.000	.000	.000	.000
VE	.733	.386	.020	.179	.530	.000	.000	.000	.000	.000
PSYM	.448	.185	.356	-.007	-.042	.798	.000	.000	.000	.000
AS	.544	.354	.157	-.141	.027	.114	.721	.000	.000	.000
MK	.836	.133	.071	.088	.044	-.035	-.103	.506	.000	.000
PSER	-.132	-.000	.021	-.084	-.048	.097	.063	-.030	.979	.000
DEPM	.095	-.048	-.011	.103	.012	-.066	-.094	.084	.064	.976

18 JOBS; 9 JOBS WITH, 9 JOBS WITHOUT HANDS-ON TESTING

	AR	SPAT	GS	CS	AS	DEPM	CPAC	PSYM	MO	MK
AR	1.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
SPAT	.730	.683	.000	.000	.000	.000	.000	.000	.000	.000
GS	.726	.208	.656	.000	.000	.000	.000	.000	.000	.000
CS	.522	.183	.056	.831	.000	.000	.000	.000	.000	.000
AS	.544	.256	.290	-.141	.731	.000	.000	.000	.000	.000
DEPM	.095	-.025	-.042	.103	-.103	.984	.000	.000	.000	.000
CPAC	.353	.196	.033	.126	-.025	.054	.904	.000	.000	.000
PSYM	.448	.395	.069	-.007	.123	-.041	.028	.788	.000	.000
MO	.635	.104	.068	.424	-.015	.010	.008	.016	.634	.000
MK	.836	.108	.105	.088	-.105	.045	.016	-.019	.067	.502

III. THE "FULL LEAST SQUARES" PREDICTOR BATTERY

TABLE F-4: THE "FULL LEAST SQUARE" BATTERY

Code	Predictor
GS	General Science
AR	Arithmetic Reasoning
NO	Numerical Operations
CS	Coding Speed
AS	Auto Shop Information
MK	Mathematical Knowledge
MC	Mechanical Comprehension
EI	Electronics Information
VE	Verbal
SPAT	Spatial composite
CPAC	Complex perceptual accuracy composite
PSYM	Psychomotor composite
SRAC	Simple reaction accuracy composite
AUTO	Autonomy composite
DEPN	Dependability composite
COND	Physical condition composite
COMB	Combat interest composite
PSER	Protective service interest composite
TECH	Technical interest composite
MACH	Machinery interest composite